

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

2M protein - nucleic search, using frame_plus_p2n model
Run on: March 1, 2004, 10:14:44 ; Search time 1716 Seconds
(without alignments)
404,131 Million cell updates/sec

Title: US-09-737-297-3
Perfect score: 79
Sequence: 1 AEGSTXDVTQNIQYAG 16

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Zgapop 6.0, Zgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DBV=slp
-Q=/cgn2/1/USPTO.spool_p/US09737297/runat_01032004_085323_3458/app_query.fasta 1.199
-DB=GenEmbl -QFMT=faetap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=150
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=50 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US09737297 @CGN 1 1 5265 @runat_01032004_085323_3458 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.cm.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	56	70.9	226486	2	AC127994	AC127994 Rattus no
2	49	62.0	174269	2	AC132961	AC132961 Rattus no
3	49	62.0	208887	2	AC134007	AC134007 Rattus no
4	49	62.0	259970	2	AC106985	AC106985 Rattus no
5	48	60.8	949	5	AF546113	AF546113 Zingel zi
6	48	60.8	207305	2	BX323822	BX323822 Danio rer
7	48	60.8	265987	2	AC094191	AC094191 Rattus no
8	47	59.5	167077	2	AC132322	AC132322 Mus muscu
9	47	59.5	167285	2	AC132853	AC132853 Mus muscu
10	47	59.5	216627	10	AL929280	AL929280 Mouse DNA
11	47	59.5	241232	2	AC120869	AC120869 Mus muscu
12	47	59.5	260424	2	AC131745	AC131745 Mus muscu
13	47	59.5	307820	2	AC130831	AC130831 Mus muscu
14	46	58.2	34753	9	AL603756	AL603756 Human DNA
15	46	58.2	146753	9	AC105753	AC105753 Homo sapi
16	46	58.2	175166	2	AC128626	AC128626 Rattus no
17	46	58.2	181443	9	AC093775	AC093775 Homo sapi
18	46	58.2	227450	2	AC103214	AC103214 Rattus no
19	45	57.0	67644	2	AC100474	AC100474 Mus muscu
20	45	57.0	110000	2	AC132794	Continuation (6 of
21	45	57.0	175770	9	AC026523	AC026523 Homo sapi
22	45	57.0	175885	2	AC113472	AC113472 Mus muscu
23	45	57.0	181884	2	AC121089	AC121089 Mus muscu
24	45	57.0	190721	8	AC120539	AC120539 Oryza sat
25	45	57.0	199645	2	BX470167	BX470167 Danio rer
26	45	57.0	225043	2	AC105657	AC105657 Rattus no
27	45	57.0	241891	2	AC134675	AC134675 Rattus no
28	45	57.0	243166	2	AC106468	AC106468 Rattus no
29	45	57.0	245002	10	AC113276	AC113276 Mus muscu
30	45	57.0	245312	2	AC113273	AC113273 Mus muscu
31	45	57.0	251495	2	AC113722	AC113722 Rattus no
32	45	57.0	259329	2	AC109570	AC109570 Rattus no
33	45	57.0	302214	2	AC117305	AC117305 Rattus no
34	44	56.3	1767	14	COMVJUM7	X00390 Mouse Hepat
35	44	55.7	947	5	AF546105	AF546105 Perca flu
36	44	55.7	32797	9	CEW01F3	292815 Caenorhabdi
37	44	55.7	36059	9	HSTCRBA120	AF009661 Homo sapi
38	44	55.7	76114	9	AL445187	AL445187 Human DNA
39	44	55.7	77743	9	HSTCRBV	U03115 Human V bet
40	44	55.7	110000	2	AC096483	Continuation (3 of
41	44	55.7	110000	2	AC112592	Continuation (2 of
42	44	55.7	136434	2	AC067868	AC067868 Homo sapi
43	44	55.7	136975	9	HSTCRB75A	AF009663 Homo sapi
44	44	55.7	152036	2	AL161776	AL161776 Homo sapi
45	44	55.7	159314	2	AC134321	AC134321 Felis cat
46	44	55.7	168958	10	AL591712	AL591712 Mouse DNA
47	44	55.7	170532	9	AL356421	AL356421 Human DNA
48	44	55.7	173385	2	AL929541	AL929541 Danio rer
49	44	55.7	215422	9	U66060	U66060 Human Germi
50	44	55.7	224429	2	AC112452	AC112452 Rattus no

51	44	55.7	233970	2	AC111904	AC111904 Rattus no	C 124	42	53.2	130454	8	AP006459	AP006459 Oryza sat
52	44	55.7	289554	2	AC125595	Rattus no	C 125	42	53.2	140028	9	AC104558	AC104558 Homo sapi
53	44	55.7	347845	2	AC096247	Rattus no	C 126	42	53.2	143354	2	AC026820	AC026820 Homo sapi
54	43.5	55.1	93629	2	AC142017	Rattus no	C 127	42	53.2	146286	9	AC027141	AC027141 Homo sapi
55	43.5	55.1	95961	2	AC130796	Rattus no	C 128	42	53.2	149806	10	AC105159	AC105159 Mus muscu
56	43.5	55.1	316230	2	AC125837	Rattus no	C 129	42	53.2	152009	2	AC010802	AC010802 Homo sapi
57	44	54.4	1899	8	AB023464	Arabidops	C 130	42	53.2	152406	10	AC122875	AC122875 Mus muscu
58	43	54.4	2040	1	CE0277601	Clostridi	C 131	42	53.2	161754	9	AC093206	AC093206 Homo sapi
59	43	54.4	10402	1	AB006303	Lactococc	C 132	42	53.2	161759	9	AL161734	AL161734 Human DNA
60	43	54.4	14364	1	AF498415	Caenorhab	C 133	42	53.2	169183	2	EX649370	EX649370 Danio rer
61	43	54.4	24534	1	AC084592	Pseudomon	C 134	42	53.2	170462	2	EX005193	EX005193 Homo sapi
62	43	54.4	24949	1	AF498403	Pseudomon	C 135	42	53.2	172525	9	AC008514	AC008514 Homo sapi
63	43	54.4	44751	8	AF498403	Pseudomon	C 136	42	53.2	172525	9	AC008514	AC008514 Homo sapi
64	43	54.4	77793	8	ATTC30C3	Arabidops	C 137	42	53.2	173320	9	AC068787	AC068787 Homo sapi
65	43	54.4	80386	8	ATTC30C3	Arabidops	C 138	42	53.2	173320	9	AC068787	AC068787 Homo sapi
66	43	54.4	106420	9	AC002546	Homo sapi	C 139	42	53.2	176130	2	AC134560	AC134560 Homo sapi
67	43	54.4	112782	2	AL3359829	Homo sapi	C 140	42	53.2	179881	2	AC074173	AC074173 Mus muscu
68	43	54.4	138989	2	AC027186	Homo sapi	C 141	42	53.2	180129	2	AC006461	AC006461 Homo sapi
69	43	54.4	143538	2	AC026555	Homo sapi	C 142	42	53.2	184150	2	AC147367	AC147367 Mus muscu
70	43	54.4	156527	9	AL161439	Human DNA	C 143	42	53.2	184485	2	AL445525	AL445525 Mus muscu
71	43	54.4	159767	2	AC080178	Homo sapi	C 144	42	53.2	184800	2	AC074144	AC074144 Mus muscu
72	43	54.4	162632	9	AC093916	Homo sapi	C 145	42	53.2	187960	9	AP000866	AP000866 Homo sapi
73	43	54.4	164176	2	AC080147	Homo sapi	C 146	42	53.2	196793	2	AC111935	AC111935 Rattus no
74	43	54.4	175448	2	AC010939	Homo sapi	C 147	42	53.2	196793	2	AC111935	AC111935 Rattus no
75	43	54.4	175668	9	CNS05TEW	Human chr	C 148	42	53.2	196824	10	AC096623	AC096623 Mus muscu
76	43	54.4	189694	5	EX072535	Zebrafish	C 149	42	53.2	198531	2	AC137379	AC137379 Rattus no
77	43	54.4	198777	8	ATCHRIV63	Human chr	C 150	42	53.2	203390	2	AC119825	AC119825 Mus muscu
78	43	54.4	224764	2	AC127660	Rattus no		42	53.2	208333	2	AC112848	AC112848 Rattus no
79	43	54.4	224764	2	AC127660	Rattus no		42	53.2	208333	2	AC112848	AC112848 Rattus no
80	43	54.4	232417	2	AC144603	Bos tauru		42	53.2	208333	2	AC112848	AC112848 Rattus no
81	43	54.4	232598	2	AC135146	Rattus no		42	53.2	208333	2	AC112848	AC112848 Rattus no
82	43	54.4	23765	2	AC126720	Rattus no		42	53.2	208333	2	AC112848	AC112848 Rattus no
83	43	54.4	254708	2	AC111761	Rattus no		42	53.2	208333	2	AC112848	AC112848 Rattus no
84	42.5	53.8	4109	1	CA0308231	-AJ308231 Coryneb		42	53.2	208333	2	AC112848	AC112848 Rattus no
85	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
86	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
87	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
88	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
89	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
90	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
91	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
92	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
93	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
94	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
95	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
96	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
97	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
98	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
99	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
100	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
101	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
102	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
103	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
104	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
105	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
106	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
107	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
108	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
109	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
110	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
111	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
112	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
113	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
114	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
115	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
116	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
117	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
118	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
119	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
120	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
121	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
122	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no
123	42	53.2	147	6	B01360	Part of cDN		42	53.2	208333	2	AC112848	AC112848 Rattus no

ALIGNMENTS

RESULT 1

AC127994

LOCUS

AC127994

DEFINITION

Rattus norvegicus clone CH230-400N13, WORKING DRAFT SEQUENCE.

ACCESSION

AC127994

VERSION

AC127994.3 GI:25074619

KEYWORDS

HTG, HTGS, PHASE2, HTGS, DRAFT, HTGS, FULLTOP.

SOURCE

Rattus norvegicus (Norway rat)

ORGANISM

Rattus

REFERENCE

1 (bases 1 to 226486)

Wuzny, D. Marie, Metzker, M. Lee, Abranzone, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, J., Ceasar, H., Chen, Z., Chu, J.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Davidson, M., Davis, C., Davy-Carroil, L., De Anda, C., Dederich, D.,

Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Georgievski, E., Geor, K., Gill, R., Grady, M., Guerra, W., Guervara, W.,

Gunnarsson, P., Hasland, W., Hamil, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Herrandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hughes, M.,

Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, R., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kowals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

Lorenshewa, L., Lounsbury, H., Lozano, R. J., Lu, X., Ma, J.,

Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A.,

Mangum, S., Mapua, P., Martin, K., Martin, R., Martinez, S.,

Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,

Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwakeleneh, O., Okwunodu, G., Olarnunsgoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C.,
 Plopper, P., Poindexter, A., Popovic, D., Primus, E., Pu, L.L.,
 Puazo, M., Quirz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rokey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shateman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Ugamai, K.,
 Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G. and Gibbs, R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

* provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 174269: contig of 174269 bp in length.

FEATURES

source

1. 174269
 /organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clones="CH230-420022"

1. 1701

/note="wgs contig"

/complement(169885..170863)

/note="clone_boundary"

clone_end:77

site:

end_sequence:BZ263863"

172459..174269

/note="wgs end extension"

clone_end:77"

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

ALIGNMENT SCORES

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-737-297-3 (1-16) x AC132961 (1-174269)

Qy

Db

RESULT 3

AC134007

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, A.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyte, M., Cree, A., D'Souza, L.,

Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Guevara, W.,

Gregory, E., Geer, K., Gill, R., Grady, M., Grady, W., Hamilton, K.,

Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hartman, R., Hines, S., Hladun, S.L., Hodgson, A., Hoque, M.,

Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshuwa, L., Louisedge, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartre, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Pioppo, F., Poindestre, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rokey, T., Rojase, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, J., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, J., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 259970)
Worley, K.C.

Direct Submission
Submitted (14-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 259970)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (22-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Sep 22, 2002 this sequence version replaced gi:21737418.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPCK
Center clone name: CH230-211120
----- Summary Statistics
Assembly program: Phrap; version 0.90329
Consensus quality: 20062 bases at least Q40
Consensus quality: 204771 bases at least Q30
Consensus quality: 208051 bases at least Q20
Estimated insert size: 221279; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: this is a 'working draft' sequence. It currently
consists of 5 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 167645: contig of 167645 bp in length
167646 167745: gap of unknown length
167746 172063: contig of 4318 bp in length
172064 172163: gap of unknown length
172164 256361: contig of 84198 bp in length
256362 256461: gap of unknown length
256462 257963: contig of 1502 bp in length
257964 259063: gap of unknown length
259064 259970: contig of 1907 bp in length.

Location/Qualifiers
1..259970
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-211120"
1..1101
/note="wgs_end extension
clone_end:T7"
3653..4479
/note="clone_boundary
clone_end:T7
site:EcoRI
end sequence:BH349734"
127886..129128
/note="wgs_contig"
169287..172063
/note="wgs_contig"

misc_feature
misc_feature
misc_feature
misc_feature

ORIGIN
Alignment Scores:
Pred. No.: 1.67e+03 Length: 259970
Score: 49.00 Matches: 7
Percent Similarity: 85.71% Conservatives: 5
Best Local Similarity: 50.00% Mismatches: 2
Query Match: 62.03% Indels: 0
DB: 2 Gaps: 0

US-09-737-297-3 (1-16) x AC106985 (1-259970)

Cy 2 CluglySerThr**AspValtyrGlnAsnileGlnTyrAla 15
Db 9906 GAGGGAAGTCTCTAGATATATTTCTTAATGTCAGTACAGT 9947

RESULT 5
AF546113
LOCUS
Zingel zingel small subunit ribosomal RNA gene, partial sequence;
Mitochondrial gene for mitochondrial product.
AF546113 949 bp DNA linear VRT 30-SEP-2003

ACCESSION
AF546113.1 GI:33333821
VERSION
AF546113.1
KEYWORDS
mitochondrion Zingel zingel
SOURCE
Zingel zingel
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
Percoidae; Percidae; Zingel.
REFERENCE
1 (bases 1 to 949)
Moyer, G.R., Sloss, B.L., Krajewski, C. and Billington, N.
AUTHORS
Phylogenetic relationships among Perca species (Teleostei:
Percidae) inferred from mtDNA sequence data
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 949)
Moyer, G.R., Sloss, B.L., Krajewski, C. and Billington, N.
AUTHORS
Direct Submission
TITLE
Submitted (12-SEP-2002) Zoology, Southern Illinois University, Life
JOURNAL
Science II, Carbondale, IL 62901, USA
FEATURES
Location/Qualifiers
1..949
/organism="Zingel zingel"

```

/organism="mitochondrion"
/mol_type="genomic DNA"
/db_xref="taxon:210855"
<!--949
/product="small subunit ribosomal RNA"

rRNA

ORIGIN
Alignment Scores:
  Pred. No.:      4.83      Length:      949
  Score:          48.00     Matches:      9
  Percent Similarity: 66.67%  Conservative: 1
  Best Local Similarity: 60.00%  Mismatches:  5
  Query Match:      60.76%   Indels:      0
  DB:                5       Gaps:      0

US-09-737-297-3 (1-16) x AF546113 (1-949)

Qy  2 GluGlySerThr***AspValTyGlnAsnIleGlnTyAlaGly 16
    |||||
Db  688 GAACGAAACATTAGATATCTACGTAAACATTCATACCTGGG 732

RESULT 6
BX323822/c
LOCUS      BX323822      207305 bp      DNA      linear      HTG 26-MAY-2003
DEFINITION Danio rerio clone DKEY-46C2, WORKING DRAFT SEQUENCE, 4 unordered
            pieces.
ACCESSION  BX323822
VERSION    BX323822.3 GI:31076124
KEYWORDS   HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE     Danio rerio (zebrafish)
ORGANISM   Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
            1 (bases 1 to 207305)
McLaren, S.
Direct Submission
Submitted (25-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfsh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On May 26, 2003 this sequence version replaced gi:29893020.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfsh-help@sanger.ac.uk
----- Project Information
Center project name: zK46C2
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 206142 bases at least Q40
Consensus quality: 206329 bases at least Q30
Consensus quality: 206599 bases at least Q20
Insert size: 207005; sum-of-contigs
Insert size: 204932; 5.3% error; agarose-fp
Quality coverage: 8.80x in Q20 bases; sum-of-contigs Quality
coverage: 8.93x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 29558: contig of 29558 bp in length
* 29559: gap of 100 bp
* 29659: contig of 117145 bp in length
* 146804: contig of 100 bp
* 146904: gap of 100 bp
* 166004: contig of 19697 bp in length
* 166601: gap of 100 bp

```

```

* 166701 207305: contig of 40605 bp in length.
FEATURES
  source
    Location/Qualifiers
      1..207305
        /organism="Danio rerio"
        /mol_type="genomic DNA"
        /db_xref="taxon:7955"
        /clone="DKEY-46C2"
        /clone_lib="DanioKey"
        1..29558
          /note="assembly fragment:00006"
          fragment_chain:1
        29659..146803
          /note="assembly fragment:01999"
          fragment_chain:1
        146904..166600
          /note="assembly fragment:01479"
          fragment_chain:1
        166701..207305
          /note="assembly fragment:01882.0"

ORIGIN
Alignment Scores:
  Pred. No.:      2.06e+03      Length:      207305
  Score:          48.00     Matches:      9
  Percent Similarity: 83.33%  Conservative: 1
  Best Local Similarity: 75.00%  Mismatches:  2
  Query Match:      60.76%   Indels:      0
  DB:                2       Gaps:      0

US-09-737-297-3 (1-16) x BX323822 (1-207305)

Qy  2 GluGlySerThr***AspValTyGlnAsnIleGln 13
    |||||
Db  176165 GAGGCGAGTGTCTGTGATGCTATGAAATATTCAA 176130

RESULT 7
AC094191/c
LOCUS      AC094191      265987 bp      DNA      linear      HTG 09-MAY-2003
DEFINITION Rattus norvegicus clone CH230-2D4, *** SEQUENCING IN PROGRESS ***
            6 unordered pieces.
ACCESSION  AC094191
VERSION    AC094191.6 GI:30467725
KEYWORDS   HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
            1 (bases 1 to 265987)
Muzny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cres, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hoques, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, L., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

```

Lorensushewa, L., Loulsegged, H., Lozado, R.J., Lu, X., Ma, J.,
 Maneshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mapua, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
 Nwackeleme, O., Okwodu, G., Olamunsgoon, A., Pal, S., Parks, K.,
 Paternek, S., Paul, H., Perez, A., Perez, L., Primus, E., Pu, L.-L.,
 Plummer, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, M., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, J., Sitter, C.D., Smajls, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorreller, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
 Taylor, R., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,
 Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstein, G., and Gibbs, R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 265987)
 Rat Genome Sequencing Consortium.

Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

On May 9, 2003 this sequence version replaced gi:24819519.

The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GAEE
 Center clone name: CH230-2D4
 Assembly program: Atlas;
 Consensus quality: 242549 bases at least Q40
 Consensus quality: 245907 bases at least Q30
 Consensus quality: 247886 bases at least Q20
 Estimated insert size: 254679; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
 * NOTE: This sequence may represent more than one clone.
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 6 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 258951: contig of 258951 bp in length
 * 258952 259051: gap of unknown length
 * 259052 260214: contig of 1163 bp in length
 * 260215 260314: gap of unknown length
 * 260315 261327: contig of 1013 bp in length
 * 261328 261427: gap of unknown length
 * 261428 262651: contig of 1224 bp in length
 * 262652 262751: gap of unknown length
 * 262752 263803: contig of 1052 bp in length
 * 263804 263904: gap of unknown length
 * 263904 265987: contig of 2084 bp in length.

FEATURES

source
 1..265987
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-2D4"
 1..1923
 /note="wgs contig"
 39271..40835
 /note="wgs_contig"

misc_feature

misc_feature

ORIGIN

Alignment Scores:
 Pred. No.: 2,73e+03 Length: 265987
 Score: 48.00 Matches: 9
 Percent Similarity: 66.6% Conservative: 1
 Best Local Similarity: 60.0% Mismatches: 5
 Query Match: 60.76% Indels: 0
 DB: 2 Gaps: 0

US-09-737-297-3 (1-16) x AC094191 (1-265987)

Qy 2 GluglySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16

Db 140233 GAAGCATCAATCATGATATCTACTTACGATACAGTACGGGGT 140189

RESULT 8

AC132322/c

LOCUS AC132322 167077 bp DNA linear HTG 03-SEP-2002
 DEFINITION Mus musculus chromosome UNK clone RP24-262F3, WORKING DRAFT
 SEQUENCE, 7 unordered pieces.

ACCESSION AC132322

VERSION AC132322.1 GI:22657783

KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE

1 (bases 1 to 167077)

AUTHORS

McPherson, J.D. and Waterston, R.H.

TITLE

The sequence of Mus musculus clone

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 167077)

AUTHORS

McPherson, J.D. and Waterston, R.H.

TITLE

Direct Submission

JOURNAL

Submitted (03-SEP-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA

COMMENT

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 Contact: submissions@wustl.edu
 ----- Project Information -----
 Center project name: M_BB0262F03

```

----- Summary Statistics -----
Sequencing vector: M15; 0%
Chemistry: Dye-terminator ET; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 163598 bases at least Q40
Consensus quality: 164279 bases at least Q30
Consensus quality: 164701 bases at least Q20
Insert size: 165000; agarose-fp
Insert size: 166477; sum-of-contigs
Quality coverage: 10.43 in Q20 bases; agarose-fp
Quality coverage: 9.61 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*
*      1          6559: contig of 6559 bp in length
*      6660       6659: gap of unknown length
*      6660       15026: contig of 8367 bp in length
*      15027       15126: gap of unknown length
*      15127       23802: contig of 8676 bp in length
*      23803       23902: gap of unknown length
*      23903       44261: contig of 20359 bp in length
*      44262       44362: gap of unknown length
*      44362       69037: contig of 24676 bp in length
*      69038       69137: gap of unknown length
*      69138       116336: contig of 47199 bp in length
*      116337     116436: gap of unknown length
*      116437     167077: contig of 50641 bp in length.
*
FEATURES             Location/Qualifiers
    1..167077
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10090"
        /chromosome="UNK"
        /clone="RP24-262F3"
misc_feature         1..6559
                    /note="assembly_name:Contig3"
misc_feature         6660..15026
                    /note="assembly_name:Contig4"
misc_feature         15127..23802
                    /note="assembly_name:Contig5"
misc_feature         23903..44261
                    /note="assembly_name:Contig6"
misc_feature         44362..69037
                    /note="assembly_name:Contig7"
misc_feature         69138..116336
                    /note="assembly_name:Contig8"
misc_feature         116437..167077
                    /note="assembly_name:Contig9"
ORIGIN
Alignment Scores:
Pred. No.:          2.57e+03           Length:          167077
Score:              47.00            Matches:          7
Percent Similarity: 85.71%            Conservative:      5
Best Local Similarity: 50.00%          Mismatches:       2
Query Match:        59.49%            Indels:           0
DB:                  2                Gaps:             0
US-09-737-297-3 (1-16) x AC132322 (1-167077)

QY      1 AlaGluGlySerThr***AspValTrGlnAsnIleGlnTyr 14
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      3419 GCAGATGCACACGGTTCTCCATCTTCAGGAATTACGAT 3378

```

AC132853/3
LOCUS 167285 bp linear HTG 18-MAR-2003
DEFINITION Mus musculus clone RP24-566E15, WORKING DRAFT SEQUENCE, 3 unordered
pieces.
AC132853 GI:29029319
AC132853.3 HTG; HTGS PHASE1; HTGS DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (base 1 to 167285)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP24-566E15
Unpublished
2 (bases 1 to 167285)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
Camarata,J., Chang,J., Chararo,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylox,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Neill,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tefsey,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (04-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 167285)
Birren,B., Nusbaum,C., Lander,E., Abouelail,A., Allen,N.,
Anderson,S., Aratchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nuyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tefsey,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (18-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 18, 2003 this sequence version replaced gi:28201729.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contract: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: 126729
Center clone name: 566 E 15

Submitted (29-NOV-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humgury@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 29, 2002 this sequence version replaced gi:25168777.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center -----
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humgury@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/projects/c_elegans/wormpep RP23-67M20 is from the RP21-23 Mouse PAC library.

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6.

Location/Qualifiers
1. 216627
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"

```

/chrOsonome="11"
/clone="RP23-67M20"
/clone_lib="RPC1-23"

cores:          3.45e+03          216627
               47.00           7
latrity:        83.33%          Conservative: 3
similarity:     58.33%          Mismatches: 2
               59.49%          Indels: 0
               10             Gaps: 0

7-3 (1-16) x AL929280 (1-216627)

```

```

3  GlySerThr****AspValTyrGlnAsnIleGlnTyr 14
|||||
3  GGTGACATTTAGACATCTACAGAAACCTTCAGTAT 104728
|||||
AC1210869          241232 bp   DNA   linear   HTG 15-DEC-2003
Mus musculus chromosome 9 clone RP23-128A11 map 9, *** SEQUENCING
IN PROGRESS ***, 7 unordered pieces.
AC120869
AC120869.4  GI:39841475
HTG; HTGS PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 241232)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus chromosome 9, clone RP23-128A11
Unpublished

2 (bases 1 to 241232)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgater, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hages, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LeRoque, K., Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Ratta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, D., Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (13-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 241232)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgater, B., Camarata, J., Chang, J., Choepel, Y., Collamore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hages, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C., Meldrum, P., Major, J., Manning, J., Matthews, C., McCarthy, M., McEwan, P., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Ratta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (15-DEC-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 15, 2003 this sequence version replaced gi:29164614.
All repeats were identified using RepeatMasker.
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RN/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L22416
Center clone name: 128_A_11

* NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence.

* as soon as it is available and the accession number will be preserved.

1 82003: contig of 82003 bp in length
82004 82103: gap of 100 bp
145811: contig of 64708 bp in length
145812 145911: gap of 100 bp
198431 198531: contig of 1520 bp in length
198532 202399: contig of 100 bp
202400 202499: contig of 100 bp
202500 22145: contig of 19646 bp in length
222145 22245: gap of 100 bp
22245 235764: contig of 13519 bp in length
235765 235864: gap of 100 bp
241232: contig of 5368 bp in length.

FEATURES

Location/Qualifiers
1..241232
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="9"
/map="9"
/clone="RP23-128A11"
/clone_lib="RPC1-23 Female Mouse BAC"

ORIGIN

Alignment Scores:
Pred. No.: 3.89e+03 Length: 241232
Score: 47.00
Percent Similarity: 71.43%
Best Local Similarity: 57.14%
Query Match: 59.49%
DB: 2 Gaps: 0

US-09-737-297-3 (1-16) x AC120869 (1-241232)
QY 3 GlySerThr***AspValTyGlnAsnIleGlnTyzAlaGly 16
DB 206096 GCGCAGCGCAGGACACCTTCTACCTACATATGCTGCG 206137

RESULT 12

AC131745/2 260424 bp DNA linear HTG 01-SEP-2002
AC131745/2 260424 bp DNA linear HTG 01-SEP-2002
LOCUS Mus musculus chromosome UNK clone RP23-5513, WORKING DRAFT
DEFINITION SEQUENCE, 9 unordered pieces.
AC131745
VERSION AC131745.2 GI:22539404
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 260424)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 260424)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT On Aug 29, 2002 this sequence version replaced gi:22476025.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.edu

```

----- Project Information -----
Center project name: M_BA005SL03
----- Summary Statistics -----
Sequencing vector: plasmid, 100%
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 256294 bases at least Q40
Consensus quality: 256803 bases at least Q30
Consensus quality: 257073 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1238: contig of 1238 bp in length
* 1239 1338: gap of unknown length
* 1339 3609: contig of 2271 bp in length
* 3610 3709: gap of unknown length
* 3710 11206: contig of 7497 bp in length
* 11207 11306: gap of unknown length
* 11307 20837: contig of 9531 bp in length
* 20838 20937: gap of unknown length
* 20938 31601: contig of 10664 bp in length
* 31602 31701: gap of unknown length
* 31702 52938: contig of 21237 bp in length
* 52939 53038: gap of unknown length
* 53039 105314: contig of 52276 bp in length
* 105315 105415: gap of unknown length
* 105415 175037: contig of 69622 bp in length
* 175037 175137: gap of unknown length
* 175137 260424: contig of 85288 bp in length.

FEATURES
Location/Qualifiers
1..260424
    /organism="Mus musculus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10090"
    /chromosome="UNK"
    /clones="RP23-55L3"
1..1238
    /note="assembly_name:Contig16"
1339..3609
    /note="assembly_name:Contig17"
3710..11206
    /note="assembly_name:Contig18"
11307..20837
    /note="assembly_name:Contig19
clone_end:T7
vector_side:right"
20938..31601
    /note="assembly_name:Contig20"
31702..52938
    /note="assembly_name:Contig21"
53039..105314
    /note="assembly_name:Contig22"
105415..175036
    /note="assembly_name:Contig23"
175137..260424
    /note="assembly_name:Contig24"

ORIGIN
Alignment Scores:
Pred. No.: 4.24e+03 Length: 260424
Score: 47.00 Matches: 8
Percent Similarity: 84.62% Conservative: 3
Best Local Similarity: 61.54% Mismatches: 2
Query Match: 59.49% Indels: 0

```

```

DB:
US-09-737-297-3 (1-16) x AC131745 (1-260424) Gaps: 0
Qy 1 AlagluglyserThr***AspValtyrGlnAsnIleGln 13
Db 73948 GCTGTGGTCCACACACAGACATTTCCAGATGCCAG 73910

RESULT 13
AC130831/c
LOCUS
DEFINITION
AC130831
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AC130831
Mus musculus chromosome UNK clone RP24-80P9, WORKING DRAFT
SEQUENCE, 15 unordered pieces.
AC130831
AC130831.2 GI:23915596
HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
Mus musculus (house mouse)
Mus musculus
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 307820)
McPherson,J.D. and Waterston,R.H.
Direct submission
Submitted (14-AUG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 307820)
McPherson,J.D. and Waterston,R.H.
Direct submission
Submitted (12-OCT-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Oct 12, 2002 this sequence version replaced gi:22218600.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BA0080P09
----- Summary Statistics -----
Sequencing vector: plasmid, 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 298785 bases at least Q40
Consensus quality: 299893 bases at least Q30
Consensus quality: 300749 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1238: contig of 1238 bp in length
* 1239 1338: gap of unknown length
* 1339 2591: contig of 1253 bp in length
* 2592 2691: gap of unknown length
* 2692 4814: contig of 2123 bp in length
* 4815 4914: gap of unknown length
* 4915 9002: contig of 4088 bp in length
* 9003 9103: gap of unknown length
* 9103 15512: contig of 6410 bp in length
* 15512 15612: gap of unknown length
* 15613 22375: contig of 6763 bp in length

```

*	22376	22475:	gap of	unknown	length
*	22476	22948:	contig of	7473	bp in length
*	22949	30048:	gap of	unknown	length
*	30049	41529:	contig of	11481	bp in length
*	41530	41629:	gap of	unknown	length
*	41630	59558:	contig of	17929	bp in length
*	59559	59658:	gap of	unknown	length
*	59659	77534:	contig of	17876	bp in length
*	77535	77634:	gap of	unknown	length
*	77635	93535:	contig of	15901	bp in length
*	93536	93535:	gap of	unknown	length
*	93536	120760:	contig of	27125	bp in length
*	120761	120860:	gap of	unknown	length
*	120861	148627:	contig of	27767	bp in length
*	148628	148727:	gap of	unknown	length
*	148728	186376:	contig of	37549	bp in length
*	186277	186376:	gap of	unknown	length
*	186377	307820:	contig of	121444	bp in length

FEATURES

```

1. .307820
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-80P9"
1. .1238
misc_feature
misc_feature
1339. .2591
/notes="assembly_name:Contig19"
1339. .2591
/notes="assembly_name:Contig43"
2692. .4814
misc_feature
/notes="assembly_name:Contig47"
4915. .9002
misc_feature
/notes="assembly_name:Contig48"
9103. .1512
misc_feature
/notes="assembly_name:Contig49"
15613. .22375
misc_feature
/notes="assembly_name:Contig50"
22476. .29948
misc_feature
/notes="assembly_name:Contig51"
30449. .41529
misc_feature
/notes="assembly_name:Contig52"
41630. .59558
misc_feature
/notes="assembly_name:Contig53"
59659. .77534
misc_feature
/notes="assembly_name:Contig54"
77635. .93535
misc_feature
/notes="assembly_name:Contig55"
93636. .120760
misc_feature
/notes="assembly_name:Contig56"
120861. .148627
misc_feature
/notes="assembly_name:Contig57"
148728. .186276
misc_feature
/notes="assembly_name:Contig58"
186377. .307820
misc_feature
/notes="assembly_name:Contig59"

```

ORIGIN

Alignment Scores:	5.12e+03	Length:	307820
Pred. No.:	47.00	Matches:	8
Score:	84.62	Conservative:	3
Percent Similarity:	61.54%	Mismatches:	2
Best Local Similarity:	59.49%	Indels:	0
Query Match:	2	Gaps:	0
DB:	52		

US-09-737-297-3 (1-16) x AC130831 (1-307820)

[illegible]

LOCUS AL603756 34753 bp DNA linear PRI 15-NOV-2001
DEFINITION Human DNA sequence from clone RP11-338121 on chromosome 10,
complete sequence.
ACCESSION AL603756
VERSION AL603756.8 GI:16973164
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 34753)

Chapman, U.	Direct Submission	Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humkey@sanger.ac.uk
		Clone requests: clonerequest@sanger.ac.uk
		On Nov 16, 2001 this sequence version replaced gi:15859997.
		During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

Only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/c_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr10> Rp11-338121 is from the library RPc1-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

FEATURES

```

sequence:
  Location/Qualifiers
    1..34753
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="10"
      /clone="RP11-338121"
      /clone_lib="RP11-338121"

```

ORIGIN

Alignment Scores:		
Pred. No.:	701	Length:
Score:	46.00	Matches:
Percent Similarity:	73.33%	Conservative:
Best Local Similarity:	53.33%	Mismatches:
Query Match:	58.23%	Indels:
DB:	9	Gaps:

US-09-737-297-3 (1-16) x AL603756 (1-34753)

1 AlaGluGlySerThr***AspValTyrGlnAsnIleGlnTyrAla 15
:::|||||:::|||||:::|||||:::|||||
4389 AGTGAAGCTATACACTGTAAACATCTACACTAACTCTCAATATGCC 4333

RESULT 15
 AC105753
 LOCUS
 DEFINITION
 AC105753 Homo sapiens chromosome 3 clone RP11-189A1, complete sequence.
 AC105753 AC068383
 AC105753.2 GI:19807859
 HTG.
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eumetazoa; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 146753)
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
 Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.
 Direct Submission
 Unpublished
 2 (bases 1 to 146753)
 Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
 Direct Submission
 Submitted (09-JAN-2002) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 3 (bases 1 to 146753)
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
 Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.
 Direct Submission
 Submitted (23-MAR-2002) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 On Mar 29, 2002 this sequence version replaced gi:18093014.

 Center: University of Washington Genome Center
 Center Code: UWGC
 Web site: <http://www.genome.washington.edu>
 Contact: uwgc@u.washington.edu
 Drafting Center: BCM

 Project Information
 Center project name: chr-3
 Center clone name: RP11-189A1 (bc0629)

 Summary Statistics
 Sequencing vector: unknown; 58% of reads
 Chemistry: L08752; 42% of reads
 Chemistry: Dye-terminator ET; 67% of reads
 Chemistry: Dye-terminator Big Dye; 33% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 146646 bases at least Q40
 Consensus quality: 146742 bases at least Q30
 Consensus quality: 146753 bases at least Q20
 Insert size: 146753; sum-of-contigs
 Quality coverage: 8.6x in Q20 bases; sum-of-contigs

 Overlapping Sequences:
 5': RP11-169G24 (UWGC:bc0281) AC099775, 142354-bp overlap
 3': Mapping in progress

 Sequence Quality Assessment:
 This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 GenBank flat file format but are available as part
 of this entry's ASN.1 file.

 This sequence was finished as follows unless otherwise noted:
 all regions were either double-stranded or sequenced with an
 alternate chemistry or covered by high quality data (i.e., Phred
 quality >= 30); an attempt was made to resolve all sequencing
 problems, such as compressions and repeats; all regions were
 covered by at least one plasmid subclone or more than one M13
 subclone; and the assembly was confirmed by restriction digest.

 Sequence Validation:
 This sequence has been validated by Multiple Complete Digest

fingerprinting. Comparison of the experimentally derived digest
 fragments with sequence-predicted fragments is given below.
 The electronically-digested sequence consists of both insert and
 vector, in order to accurately represent the entire circular BAC.
 Small fragments below a variable cutoff (approximately 400-800 bp)
 are not resolved in the fingerprint and hence do not appear
 in the table. There are no significant remaining discrepancies
 between the experimental and predicted values. Uniquely ordered
 fragments are separated by dashed lines.

BglII	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	HindIII
-----	-----	-----	-----	-----	-----	-----	-----
3452	3412	8854	8854	1266	1202		
2067	2069	<800	<800	6382	6434		
6750	6959	2170	2209	512	<800		
2496	2548	2229	2209	449	<800		
1613	1616	10412	10178	1573	1540		
697	<800	511	<800	2788	2829		
10754	10865	2736	2777	5427	5353		
13989	14417	1612	1596	3646	3639		
993	1010	11781	11693	1142	1202		
1336	1334	129	<800	5739	5732		
5634	5587	283	<800	3145	3062		
4377	4362	9274	9302	4847	4798		
3521	3551	1887	1813	1010	1001		
2725	2778	7494	7606	6374	6434		
4804	4727	1254	1259	9723	9778		
2022	2069	1824	1813	575	<800		
4922	4942	4502	4486	548	<800		
518	<800	10009	10178	7956	8044		
2926	2951	8857	8854	4131	4023		
7025	6959	2882	2894	6879	6899		
3536	3551	3745	3762	620	<800		
1441	1469	1030	1029	66	<800		
127	<800	849	883	6208	6156		
262	<800	6230	6194	5230	5125		
2410	2412	895	883	2261	2282		
128	<800	1793	1813	165	<800		
10968	10865	5779	5728	3668	3639		
2874	2951	1333	1269	1402	1373		
818	760	1545	1509	2990	3062		

186	<800	7591	7606	889	904
2432	2412	1709	1813	328	<800
1511	1469	3172	3180	2625	2681
11719	11520	2207	2209	1257	1202
10009	9871	8	<800	2607	2681
11172	10865	3339	3317	10360	10292
749	760	106	<800	555	<800
350	<800	5918	5728	3700	3639
4251	4241	37	<800	3629	3639
2061	2069	4762	4724	11989	11982
5767	5736	3298	3317	764	<800
66	<800	2012	2002	3033	3062
		6234	6194	2718	2681
		3131	3180	1888	1878
		306	<800	542	<800
				2236	2282
				1194	1202
				2574	2681
				172	<800
				707	<800
				1177	1202
				3792	3823

FEATURES

Location/Qualifiers
1. 146753
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-189A1"
/clone_lib="RP11-189A1"

ORIGIN

Alignment Scores:
Pred. No.: 3,548+03 Length: 146753
Score: 46.00 Matches: 8
Percent Similarity: 83.33% Conservative: 2
Best Local Similarity: 66.67% Mismatches: 2
Query Match: 58.23% Indels: 0
DB: 9 Gaps: 0

US-09-737-297-3 (1-16) x AC105753 (1-146753)

QY 4 SerThr***AspValrGlnAsnIleGlnrAla 15

Db 104726 AGCACACAGACATACACAGAACATTCAGTACAGC 104761

RESULT 16

AC128626 175166 bp DNA linear HTG 19-NOV-2002
LOCUS Rattus norvegicus clone CH230-362b20, WORKING DRAFT SEQUENCE, 2
DEFINITION unordered pieces.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC128626.3 GI:25085439
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 175166)
Munzy D.M., Metzker M., Lee S., Abramson S., Adams C., Alder J.,
Allen C., Allen H., Alsbrooks S., Amin A., Anguiano D.,
Anyalebechi V., Ayagi A., Ayodeji M., Baca E., Baden H.,
Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benahmed F.,
Biswal K., Blair J., Blankenburg K., Blyth P., Brown M.,
Bryant N., Buhay C., Burch P., Burrell K., Calderon E.,
Cardenas V., Carter K., Cavazos I., Cesar H., Center A.,
Chacko J., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,
Claveland C., Cockrell R., Cox C., Coyle M., Cree A., D'Souza L.,
Davila M.L., Davis C., Davy-Carroll L., De Anda C., Dederich D.,
Delgado O., Denison S., Deramo C., Ding Y., Dinh H., Divya K.,
Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duval B., Eaves K.,
Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,
Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,
Frazer C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,
Gregorogis E., Geer K., Gill R., Grady M., Guerra W., Guevara W.,
Gunaratne P., Haaland W., Hamill C., Hamilton C., Hamilton K.,
Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J.,
Hernandez R., Hines S., Hladum S.L., Hodgson A., Hognes M.,
Hollins B., Howells S., Hulyk S., Hume J., Idlebird D., Jackson A.,
Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jolivet A.,
Karpathy S., Kelly S., Kelly S., Khan Z., King L., Kovar C.,
Kowis C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J.,
Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J.,
Lorensuwa L., Louise H., Lozano R.J., Lu X., Ma J.,
Maheshwari M., Mahindratne M., Mahmoud M., Malloy K., Mangum A.,
Mangum B., Mapa P., Martin K., Martin R., Martinez E.,
Mawhney S., McLeod M.P., McNeill T.Z., Meenen E.,
Milosavljevic A., Miner G., Minja E., Montemayor J., Moore S.,
Morgan M., Morris K., Morris S., Muidasa M., Murphy M., Nair L.,
Nankervis C., Neal D., Newton N., Nguyen N., Norris S., Parks K.,
Nwakoileme O., Okwono G., Olarnpungoon A., Pal S., Parks K.,
Pasternak S., Paul H., Perez A., Perez L., Pfannkuch C.,
Plopper F., Poindexter A., Popovic D., Primus E., Pu L.L.,
Puro M., Quiroz J., Rachlin E., Reeves K., Regier M.A., Reigh R.,
Reilly B., Reilly M., Ren Y., Reuter M., Richards S., Riggs F.,
Rives C., Rodkey T., Rojas A., Rose M., Rose R., Ruiz S.J., Shen H.,
Sanders W., Savary G., Scherer S., Scott G., Shatsman S., Shen H.,
Shetty J., Shvartsbayn A., Sisson I., Sitter C.D., Smajs D.,
Sneed A., Sodergren E., Song X.-Z., Sorelle R., Sosa J.,
Steinle M., Strong R., Sutton A., Svatek A., Tabor P., Taylor C.,
Taylor T., Thomas N., Thomas S., Tingey A., Trejos Z., Ugmani K.,
Valas R., Vera V., Villaseca D., Waldron L., Walker B., Wang J.,
Wang Q., Wang S., Warren J., Warren R., Wei X., White P.,
Williams G., Willson R., Wleczek R., Wooden H., Worley K.,
Wright D., Wright R., Wu J., Yakub S., Yen J., Yoon L., Yoon V.,
Yu F., Zhang J., Zhou J., Zhou X., Zhao S., Dunn D., von
Niederhausern A., Weiss R., Smith D.R., Holt R.A., Smith H.O.,
Weinstock G. and Gibbs R.A.

TITLE

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 175166)
Worley K.C.

TITLE

Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

TITLE

3 (bases 1 to 175166)
Rat Genome Sequencing Consortium.

REFERENCE

Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23907895.

COMMENT

of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fingerprints with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

BgIII				BcoRI				HindIII			
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
2650	2653	8696	8688	4702	4760						
2067	2046	6	<800	6382	6418						
5116	5151	2242	2232	512	<800						
7479	7512	6880	6819	449	<800						
5124	5151	820	853	4526	4580						
1368	1389	121	<800	3415	3369						
361	<800	5850	6137	766	782						
1434	1389	8533	8531	2372	2383						
6652	7171	5619	5668	3746	3647						
7221	7171	4626	4621	2634	2633						
26	<800	2939	2886	261	<800						
1410	1389	1092	1052	1864	1863						
637	<800	367	<800	5348	5073						
1405	1389	303	<800	610	<800						
6504	6559	3131	3281	503	<800						
66	<800	6234	6137	1663	1646						
5767	5695	2012	2004	13333	13598						
2061	2046	3298	3281	3792	4024						
4251	4209	4762	4621	1177	1160						
350	<800	37	<800	707	<800						
749	743	5818	5668	172	<800						
11172	10717	106	<800	2574	2633						
10009	9846	3339	3152	1194	1160						

Alignment Scores:

Pred. No.: 4,49e+03
 Score: 46.00
 Percent Similarity: 83.33%
 Best Local Similarity: 66.67%
 Query Match: 58.23%
 DB: 9

Length:

Matches: 181443
 Conservatives: 8
 Mismatches: 2
 Indels: 0
 Gaps: 0

11719	11437	8	<800	2236	2251
1511	1501	2207	2232	542	<800
2432	2407	3172	3281	1888	1863
186	<800	1709	1809	2718	2633
818	842	7591	7579	3033	3026
2874	2933	1545	1525	764	782
10963	10717	1333	1264	11989	11774
128	<800	5779	5668	3629	3647
2410	2407	1793	1809	3700	3647
262	<800	895	853	555	<800
127	<800	6225	6137	10360	10193
1441	1389	949	853	2607	2633
3536	3533	1030	1052	1257	1220
7025	7171	3745	3747	2620	2633
2926	2933	2882	2886	328	<800
518	<800	8857	8868	889	917
4922	4919	10009	10026	2990	3026
2022	2046	4502	4408	1402	1373
4804	4707	1824	1809	3668	3647
2725	2775	1254	1264	165	<800
3521	3533	7494	7579	2261	2251
4377	4343	1887	1864	5230	5229
5634	5695	9274	9344	6208	6123
1336	1389	263	<800	66	<800
993	1002	129	<800	620	<800
13989	14546	11781	11712	6879	6897
10754	10717	1612	1670	4131	4392
697	<800	2736	2886	7956	7994
1613	1593	511	<800	548	<800
		10415	10337	575	<800
				9723	9716
				6374	6418
				1010	998

US-09-737-297-3 (1-16) x AC099775 (1-181443)

QY 4 SerThr**aspValtyrGlnAsnIleGlyTyrAla 15

Db 143815 ACCACACAGACATACAGACATTCAGTACAGC 143850

RESULT 18

AC103214/c

LOCUS

DEFINITION Rattus norvegicus clone CH230-13K21, *** linear HTG 13-MAY-2003

*** SEQUENCING IN PROGRESS

ACCESSION

AC103214.5 GI:30578638

VERSION

HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

Rattus norvegicus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 227450)

Muzny D, Marie, Metzker M, Lee, Abramson S., Amin A., Anguiano D.,

Allen C., Allen H., Alsbrooks S., Burch P., Burrell K., Calderon E.,

Baldwin D., Sandarandake D., Barber M., Barnstead M., Benahmed F.,

Biswal K., Blair J., Blankenburg K., Blyth P., Brown M.,

Bryant N., Buhay C., Burch P., Burrell K., Calderon E.,

Cadenas V., Carter K., Cavazos I., Caesar H., Center A.,

Chacko J., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,

Cleaveland C., Cockrell R., Cox C., Coyle M., Cree A., D'Souza L.,

Davila M.L., Davis C., Davy-Carroll L., De Anda C., Dederich D.,

Delgado O., Denson S., Deramo C., Ding Y., Dinh H., Divya K.,

Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duval K.,

Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,

Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,

Fraser C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,

Gebregorgis E., Geer K., Gill R., Grady M., Guerra W., Guevara W.,

Gunnarone P., Haaland W., Hamil C., Hamilton C., Hamilton K.,

Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J.,

Hernandez R., Hines S., Hladun S.L., Hodgson A., Hughes M.,

Hollins S., Howells S., Rulyk S., Rume J., Idelbird D., Jackson A.,

Johnson L., Jacobs L., Jiang H., Johnson B., Johnson R., Jolivet A.,

Karpachy S., Kelly S., Kelly S., Khan Z., King L., Kovar C.,

Kowitz C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J.,

Liu J., Liu Y., Liu Y., London P., Longacre S., Lopez J.,

Lorenshewa L., Louised H., Lozano R.J., Lu X., Ma J.,

Maheshwari M., Mahindartne M., Mahmoud M., Malloy K., Mangum A.,

Mangum B., Mapa P., Martin K., Martin R., Martinez E.,

Manniney S., McLeod M.P., McNeil I.Z., Meenen E.,

Milosavljevic A., Miner G., Minja E., Montemayor J., Moore S.,

Morgan M., Morris K., Morris S., Munidasa M., Murphy M., Nair L.,

Nankervis C., Neal D., Newton N., Nguyen N., Norris S.,

Nwaokelemeh O., Okwuonu G., Olarnpunsagoon A., Pal S., Parks K.,

Pasternak S., Paul H., Perez A., Perez L., Pfannkuch C.,

Plopper F., Poinexter A., Popovic D., Primus E., Pu L.-L.,

Puazo M., Quiroz J., Rachin E., Reeves K., Regier M.A., Reigh R.,

Reilly B., Reilly M., Ren Y., Reuter M., Richards S., Riggs P.,

Rives C., Rodkey R., Rojas A., Rose M., Rose R., Ruiz S.J.,

Sanders W., Savery G., Scher S., Scott G., Shatsman S., Shen H.,

Shetty J., Shvartsbeyn A., Sisson I., Sitter C.D., Smajls D.,

Sneed A., Sodergren E., Sung X.-Z., Sorelle R., Sosa J.,

Stearle M., Strong R., Sutton A., Svatek A., Tabor P., Taylor C.,

Taylor T., Thomas N., Thomas S., Tingey A., Trejos Z., Usmani K.,

Valas R., Vera V., Villasana D., Waldron L., Walker B., Wang J.,

Wang Q., Wang S., Warren J., Warren R., Wei X., White P.,

Williams G., Willson R., Wlezyk R., Woodden H., Worley K.,

Wright D., Wright R., Wu J., Yakub S., Yen J., Yoon L., Yoon V.,

Yu F., Zhang J., Zhou J., Zhou X., Zhao S., Dunn D., von

Niederhausern A., Weiss R., Smith D.R., Holt R.A., Smith H.O.,

Weinstock G. and Gibbs R.A.

Direct Submission

Unpublished

2 (bases 1 to 227450)

TITLE

JOURNAL

REFERENCE

Worley K.C.

Direct Submission

Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 227450)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On May 13, 2003 this sequence version replaced gi:23265300.

The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GJME

Center clone name: CH230-13K21

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 214748 bases at least Q40

Consensus quality: 217586 bases at least Q30

Consensus quality: 219134 bases at least Q20

Estimated insert size: 224175; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 1 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 227450: contig of 227450 bp in length.

----- Location/Qualifiers

FEATURES

source

1..227450

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-13K21"

misc_feature

1..2929

/note="wgs end extension

clone end:Sp6"

7740..8494

/note="clone boundary

clone end:Sp6"

site:EcoRI

end sequence:BH274984"

143836..145156

/note="wgs contig"

220747..221562

/note="clone boundary

clone end:r7

site:EcoRI

end sequence:BH274983"

misc_feature

misc_feature

Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 175770)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Bouckgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P., DeArallano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Pollara, V., Raymond, J., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Sougne, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (03-JUL-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 175770)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Bouckgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P., DeArallano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Submitted (11-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 11, 2001 this sequence version replaced gi:14589636.
All repeats were identified using RepeatMasker.
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L6959
Center clone name: 461_F_11

FEATURES

Location/Qualifiers
source
1. 175770
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-461F11"
/clone_lib="RP11-461F11 Human Male BAC"
complement(22. 147)
/rpt_family="L1M4"
complement(346. 1379)
/rpt_family="L1ME1"
complement(1606. 1975)
/rpt_family="L1M4"
complement(1976. 3813)
/rpt_family="L1PA6"
3816. 3915
/rpt_family="L1PA6"
complement(3932. 4110)
/rpt_family="L1M4"
complement(4157. 4674)
/rpt_family="L1MA3"
complement(4677. 5404)
/rpt_family="L1MA3"
5411. 5464
/rpt_family="L1MA3"
5476. 5511
/rpt_family="AT_rich"
5512. 5857
/rpt_family="L1MA3"
5873. 6002
/rpt_family="L1MA3"
6499. 6679
/rpt_family="L1M4a"
6682. 6709
/rpt_family="AT_rich"
complement(7571. 7722)
/rpt_family="MIR"
complement(7750. 8026)
/rpt_family="AluSg"
8578. 8698
/rpt_family="L1M1"
8913. 9129
/rpt_family="L1M4a"
9638. 9659
/rpt_family="AT_rich"
9838. 9986
/rpt_family="MIR3"
11426. 11539
/rpt_family="L1MEC"
11548. 11742
/rpt_family="L1MB7"
11770. 14350
/rpt_family="L1ME"
14356. 14479
/rpt_family="L1ME"
complement(14495. 14873)
/rpt_family="THE1C"
14884. 15039
/rpt_family="L1MA10"
15438. 15625
/rpt_family="L13"
15626. 16528
/rpt_family="L1PA10"
16529. 16659
/rpt_family="THE1C"
16660. 17689
/rpt_family="L1TR5"

```

repeat_region 17690..17932
/rpt_family="THEIC"
repeat_region 17933..18041
/rpt_family="L1PA10"
repeat_region 18042..18178
/rpt_family="L13"
repeat_region 18426..18694
complement(18506)
/rpt_family="MIR"
19480..19660
/rpt_family="MLT1J"
20087..20124
/rpt_family="(TAA)n"
repeat_region complement(20158)
/rpt_family="L1T854"
20743..20767
/rpt_family="AT_rich"
21711..21858
/rpt_family="MER5A"
repeat_region complement(22055)
/rpt_family="MIR"
24503..24578
/rpt_family="(TATAA)n"
24695..24760
/rpt_family="AT_rich"
repeat_region complement(25104)
/rpt_family="MIR"
25743..25815
/rpt_family="L3"

```

Alignment Scores:

Pred. No.:	6.9e+03	Length:	175770
Score:	45.00	Matches:	8
Percent Similarity:	76.92%	Conservative:	2
Best Local Similarity:	61.54%	Mismatches:	3
Query Match:	56.96%	Indels:	0
DB:	9	Gaps:	0

US-09-737-297-3 (1-16) x AC026523 (1-175770)

QY 3 GlySerThr***ApValTyGlnAsnIleGlnTyAla 15
 DB 97542 GGGACTGAAGCAACACATACCAACATACATGTCGA 97580

RESULT 22
 AC113472/c 175885 bp DNA linear HTG 16-JUL-2003
 LOCUS Mus musculus clone RP23-279A14, WORKING DRAFT SEQUENCE, 6 ordered
 DEFINITION pieces.

ACCESSION AC113472.4 GI:32813662
 VERSION HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 175885)

REFERENCE Birren, B., Nusbaum, C. and Lander, E.

AUTHORS Mus musculus, clone RP23-279A14

TITLE Unpublished

JOURNAL

REFERENCE 2 (bases 1 to 175885)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, B., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagsos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczyk, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McSwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,

Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Struse, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

REFERENCE

AUTHORS

Submitted (01-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 175885)
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagsos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Mathews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Stange-Thomann, N., Stojanovic, N., Topham, K., Travers, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (16-JUL-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 16, 2003 this sequence version replaced gi:21328538.
 All repeats were identified using RepeatMasker:
 Smit, A. F. A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE

JOURNAL

REFERENCE

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L23660
 Center clone name: 279_A.14
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 174774 bases at least Q40
 Consensus quality: 175195 bases at least Q30
 Consensus quality: 175309 bases at least Q20
 Insert size: 17600; agarose-ff
 Insert size: 17338; sum-of-ctnigs
 Quality coverage: 12.3 in Q20 bases; agarose-ff
 Quality coverage: 12.4 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 925: contig of 925 bp in length
 * 926 1025: gap of 100 bp
 * 1026 14898: contig of 13873 bp in length
 * 14899 14998: gap of 100 bp

* 14999 30359: contig of 15361 bp in length
 * 30360 30459: gap of 100 bp
 * 30460 105604: contig of 75145 bp in length
 * 105605 105704: gap of 100 bp
 * 105705 146900: contig of 41196 bp in length
 * 146901 147000: gap of 100 bp
 * 147001 175895: contig of 28885 bp in length.

FEATURES

Location/Qualifiers
 1..175885
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone="RP23-279A14"
 /clone.lib="RP23-279A14"
 /notes="assembly_fragment"
 1..925
 /notes="assembly_fragment"
 1026..14898
 /note="assembly_fragment"
 14999..30359
 /notes="assembly_fragment"
 30460..105604
 /notes="assembly_fragment"
 105705..146900
 /notes="assembly_fragment"
 147001..175885
 /note="assembly_fragment"
 clone_end:17
 vector_side:right"

ORIGIN

Alignment Scores:
 Pred. No.: 6.9e+03 Length: 175885
 Score: 45.00 Matches: 8
 Percent Similarity: 85.71% Conservative: 4
 Best Local Similarity: 57.14% Mismatches: 2
 Query Match: 56.96% Indels: 0
 DB: 2 Gaps: 0

US-09-737-297-3 (1-16) x AC113472 (1-175885)

QY 2 GlucylSerThr***AspValTyrGlnAanIleGlnTyrAla 15

Db 60842 GAGGATCAACTTTCATCTTATGGAACATCCAGTTTCT 60801

RESULT 23

AC121089/c
 LOCUS AC121089 181884 bp DNA linear HTG 16-FEB-2003
 DEFINITION Mus musculus clone RP24-16219, WORKING DRAFT SEQUENCE, 6 unordered pieces.

AC121089

AC121089.3 GI:28394949

HTG; HTGS PHASE1; HTGS DRAFT.

Mus musculus (house mouse)

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 181884)

Birken, B., Nusbaum, C. and Lander, E.

Mus musculus, clone RP24-16219

2 (bases 1 to 181884)

Birken, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,

Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,

Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A.,

Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,

Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,

Gande, S., Gord, S., Goyette, M., Graham, L., Johnson, R., Jones, C.,

Hagos, B., Horton, L., Hulme, W., Iliev, I., Janssen, R., Jones, C.,

Kamat, A., Karatas, A., Kells, C., Lakoque, K., Lamazares, R.,

Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,

MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,

McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L.,

Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,

Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,

Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,

Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,

Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,

Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,

Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,

Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

* 45725 59479: contig of 13755 bp in length
 * 59480 59579: gap of 100 bp
 * 75320: contig of 15741 bp in length
 * 75321 137186: contig of 100 bp
 * 75421 137186: contig of 61766 bp in length
 * 137187 137287: gap of 100 bp
 * 137287 181884: contig of 44598 bp in length.

FEATURES

source

Location/Qualifiers
 1. 181884
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone_lib="RP24-16219"
 /clone_lib="RPCI-24 Male Mouse BAC"
 1. 44826
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 44927..45624
 /note="assembly_fragment"
 45725..59479
 /note="assembly_fragment"
 59580..75320
 /note="assembly_fragment"
 75421..137186
 /note="assembly_fragment"
 137287..181884
 /note="assembly_fragment"
 clone_end:R7
 vector_side:right

ORIGIN

Alignment Scores:
 Pred. No.: 7.17e+03 Length: 181884
 Score: 45.00 Matches: 8
 Percent Similarity: 85.71% Conservative: 4
 Best Local Similarity: 85.71% Mismatches: 2
 Query Match: 56.96% Indels: 0
 DB: 2 Gaps: 0

US-09-737-297-3 (1-16) x AC121089 (1-181884)

QY 2 GluclySerThr***AspValTyGlnAsnIleGlnTyAla 15

Db 105401 GAAGATCAATTCATCTTATGTCACATCCAGTTTCT 105360

RESULT 24

AC120539

LOCUS 190721 bp DNA linear PLN 09-OCT-2003

DEFINITION Orzya sativa chromosome 11 BAC clone OSJNB0063D09, complete

ACCESSION

AC120539

VERSION AC120539.4 GI:37591196

KEYWORDS

HTG. Orzya sativa (japonica cultivar-group)

SOURCE

ORGANISM Orzya sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Orzyaceae; Orzya.

REFERENCE

AUTHORS

Buell,R., Hsiao,J., Ziemann,V., Moffat,K.M., Hill,J.,
 Gansberger,K., Burgess,S., Jarrari,S., Shvartsbeyn,M., Brenner,M.,
 Ciecko,A., Pai,G., Vanaken,S., Hansen,C., Utterbach,T.,
 Feidlyum,T., Khalak,H.G., Yuan,Q., Quackenbush,J., White,O.,
 Salzberg,S. and Fraser,C.

Orzya sativa ssp. japonica cv. Nipponbare OSJNB0063D09 BAC genomic

sequence

Unpublished

2 (bases 1 to 190721)

REFERENCE

AUTHORS

Buell,R.

Direct Submission

TITLE

JOURNAL

Submitted (08-MAY-2002) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA

REFERENCE

AUTHORS

Buell,R.

Direct Submission

TITLE

JOURNAL

Submitted (28-SEP-2002)

The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

REFERENCE

AUTHORS

Buell,R.

Direct Submission

TITLE

JOURNAL

Submitted (09-OCT-2003)

The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

COMMENT

On Oct 9, 2003 this sequence version replaced gi:23343711.

FEATURES

Location/Qualifiers

1. 190721

/organism="Orzya sativa (japonica cultivar-group)"

/mol_type="genomic DNA"

/cultivar="Nipponbare"

/sub_species="japonica"

/db_xref="taxon:39947"

/chromosome="11"

/clone="OSJNB0063D09"

ORIGIN

Alignment Scores:

Pred. No.: 7.56e+03 Length: 190721
 Score: 45.00 Matches: 8
 Percent Similarity: 76.92% Conservative: 2
 Best Local Similarity: 61.54% Mismatches: 3
 Query Match: 56.96% Indels: 0
 DB: 8 Gaps: 0

US-09-737-297-3 (1-16) x AC120539 (1-190721)

QY 2 GluclySerThr***AspValTyGlnAsnIleGlnTyAla 14

Db 142988 GAAGAGTACACTAGATTATTTTCATATATTAGTAT 143026

RESULT 25

EX470167

LOCUS

DEFINITION

Orzya sativa

199645 bp DNA linear

HTG 30-JUN-2003

ACCESSION

EX470167

VERSION

EX470167.4 GI:32398430

KEYWORDS

HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.

SOURCE

Danio rerio (zebrafish)

ORGANISM

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

REFERENCE

AUTHORS

McLaren,S.

Direct Submission

TITLE

JOURNAL

Submitted (28-JUN-2003)

Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

zfish-help@sanger.ac.uk

On Jul 1, 2003 this sequence version replaced gi:30424233.

COMMENT

On Jul 1, 2003 this sequence version replaced gi:30424233.

Center project name: zkp86c2

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: zfish-help@sanger.ac.uk

Center project name: zkp86c2

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 199046 bases at least Q40

Consensus quality: 199155 bases at least Q30

Consensus quality: 199307 bases at least Q20

Insert size: 199545; sum-of-contigs

Insert size: 20128; 2.0% error; agarose-fp

Quality coverage: 8.39x in Q20 bases; sum-of-contigs Quality

coverage: 8.33x in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 70421: contig of 70421 bp in length
 70422 70521: gap of 100 bp
 70522 199645: contig of 129124 bp in length.

FEATURES

Location/Qualifiers
 1. 199645
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="DKXP-86G2"
 /clone_lib="DanioKeyPilot"
 1. 70421
 /note="assembly_fragment:01572.0"
 70522. 199645
 /note="assembly_fragment:02376"

ORIGIN

Alignment Scores:
 Pred. No.: 7.96e+03 Length: 199645
 Score: 45.00 Matches: 7
 Percent Similarity: 90.00% Conservative: 2
 Best Local Similarity: 70.00% Mismatches: 1
 Query Match: 56.96% Indels: 0
 DB: 2 Gaps: 0

US-09-737-297-3 (1-16) x BX470167 (1-199645)

Qy 7 AspValTyrGlnAsnIleGlnTyrAlaGly 16

Db 25525 GACATCTATGAGACATACATACTCTGGA 25554

RESULT 26

AC105657/c
 LOCUS AC105657 225043 bp DNA linear HTG 21-SEP-2002
 DEFINITION Rattus norvegicus clone CH230-194M10, *** SEQUENCING IN PROGRESS
 *** 4 unordered pieces.
 AC105657
 AC105657.4 GI:23264257
 HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus

REFERENCE

1 (bases 1 to 225043)
 Muzny, D., Maric, Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blau, P., Blum, K., Blyth, P., Brown, M., Bryant, N., Buay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, J., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, D., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabis, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Georgegeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, N., Hognes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowals, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorusshewa, L., Louisegh, H., Lozano, R., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Muidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwankwem, O., Okwodu, G., Olarnpungoon, A., Pal, S., Pankoch, C., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, S., Pu, L., Fuazo, M., Quiroz, J., Rachin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, R., Rose, R., Ruiz, S., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Snajd, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanasa, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.
 Direct Submission
 2 (bases 1 to 225043)
 Unpublished
 Worley, K.C.
 Direct Submission
 Submitted (09-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 225043)
 Direct Submission
 Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Sep 21, 2002 this sequence version replaced gi:21736557.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence contigs will be indicated in the feature table.

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GNLM

Center clone name: CH230-194M10

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 215193 bases at least Q40

Consensus quality: 21518 bases at least Q30

Consensus quality: 215167 bases at least Q20

Estimated insert size: 230806; sum-of-contigs estimation

Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 4 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Georgeogis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, M., Gunaratne, P., Haaland, K., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Huiyk, S., Hume, J., Idelbird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovac, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenauhewa, L., Loulsegod, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmood, M., Malloy, K., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinecz, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Minosavljevic, A., Miner, G., Nanj, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundayasa, M., Murphy, M., Naif, L., Nankervis, C., Neal, D., Newdon, N., Nguyen, N., Norris, S., Nwakoileme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Pioppfer, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puazo, M., Quintero, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajz, D., Sneed, A., Sodergren, E., Song, X., Sorelle, R., Sosa, J., Streinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valae, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, K., Wel, X., White, F., Williams, G., Willson, R., Wlecsyck, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Harvey, R., Navlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Huiyik, S., Hume, J., Idiebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Khan, Z., King, L., Kovac, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenzeta, L., Loulged, H., Lozado, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mainland, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangwin, S., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mathewey, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, M.,
Nwakenleh, O., Okunnu, G., Olariupsagoon, A., Pal, S., Parks, K.,
Patel, N. S., Paul, H., Perez, A., Perez, L. H., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Prums, E., Pu, L.,
Puzo, M., Quizoz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shaerman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasanad, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.

Lorenzenewa, L., Lounsbegre, H., Lozada, R.J., Luján, Ma. J.,
 Maneshwari, M., Maindardne, M., Mamoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Matti, K., Martin, R., Martinez, E.,
 Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Mielosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwackelmech, O., Okwuonu, G., Olarunpunsagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C.,
 Platter, P., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puazo, M., Quiróz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shaumsan, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajd, S.,
 Sneed, J., Sodergren, E., Song, X.-Z., Sozalle, P., Soza, J.,
 Steidle, M., Strong, R., Suton, A., Svatek, A., Taber, P., Taylor, C.,
 Taylor, I., Thomas, N., Thomas, S., Tiggey, A., Trejos, Z., Usmani, K.,
 Valdes, K., Vera, V., Villanasa, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G. and Gibbs, R.A.

Nankervis, C., Neal, D., Newcom, G., Olariupunsagoon, A., Pal, S., Parks, K.,
Nwakolamelen, O., Okwuonu, G., Olariupunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plommer, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puzao, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, X., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savory, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajic, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valae, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, K., Wei, X., White, F.,
Williams, G., Willson, R., Wlezycki, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs R.A.

Propper, F., Poindestre, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, X., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojass, A., Rose, M., Rose, R., Ruiz, S.J., Sanders W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steckelm, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasantana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlarczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs R.A.

Sneddy, J., Snwaarsbeyn, A., Sisson, J., Sitterle, C.D., Smajcs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas S., Tingey, A., Trejos, Z., Usmani, C.,
Valas, R., Vera, V., Villasan, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlciayk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs R.A.

Wang, Q., Wang, S., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Mieczys, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs R. A.

Wiegand, D., Wiegand, K., Wu, C., Takum, S., Ren, J., Itoh, E., Itoh, V., Yu, F., Zhang, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Consensus quality: 212024 bases at least Q20
 Consensus quality: 212664 bases at least Q20
 Estimated insert size: 216088: sum-of-contigs estimation
 Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces
 is believed to be correct as given, however the sizes
 of the gaps between them are based on estimates that have
 been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

* 1 241891: contig of 241891 bp in length.

FEATURES

source
 1. 241891
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-116C9"
 1369..2244
 /notes="clone boundary
 clone end:Sp6
 site:EcoRI
 end sequence: BH322284"
 misc_feature
 complement(238654..239483)
 /notes="clone boundary
 clone end:T7
 site:EcoRI
 end sequence: BH322283"

ORIGIN

Alignment Scores:
 Pred. No.: 9.88e+03 Length: 241891
 Score: 45.00 Matches: 9
 Percent Similarity: 76.92% Conservative: 1
 Best Local Similarity: 69.23% Mismatches: 3
 Query Match: 56.96% Indels: 0
 DB: 2 Gaps: 0

US-09-737-297-3 (1-16) x AC132675 (1-241891)

Qy 4 SerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16

Db 23704 AGCACTACTGATCAATCAACATCAATCTCTGGG 23742

RESULT 28
 AC106468/c
 LOCUS AC106468 243166 bp DNA linear HTG 13-MAY-2003
 DEFINITION Rattus norvegicus clone CH230-212C19, WORKING DRAFT SEQUENCE, 5
 unorderd pieces.

ACCESSION AC106468

VERSION AC106468.4 GI:30580929

KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS FULLTOP.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 243166)

REFERENCE
 AUTHORS Muzny, D., Marie, M., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, B., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Guevara, P., Geer, K., Gill, R., Grady, M., Guerra, W., Guera, W., Gunaratne, S., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, J., London, P., Longacre, S., Lopez, J., Lorenshuwa, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, S., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okwuonu, G., Olarinpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfanckoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shivartsbeyn, A., Sisson, I., Sitter, C.D., Sosa, J.D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.D., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

Center: Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GKYT

Center clone name: 383_E_21

```

FEATURES
  source
    Location/Qualifiers
      1..245002
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10090"
        /chromosome="5"
        /map="5"
        /clone="RP23-383E21"
        /clone_lib="RPCI-23 Female Mouse BAC"
      1..6
        misc_feature
          note="clone boundary"
          clone_end:SP6
          site:EcoRI
        repeat_region
          complement(272..517)
          /rpt_family="B4A"
        repeat_region
          837..1001
          /rpt_family="RSINE1"
        repeat_region
          complement(1239..1384)
          /rpt_family="B1_MM"
        repeat_region
          1486..1649
          /rpt_family="MER3"
        repeat_region
          complement(1759..1888)
          /rpt_family="B1F"
        repeat_region
          2501..2566
          /rpt_family="(TA)n"
        repeat_region
          2711..2731
          /rpt_family="AT_rich"
        repeat_region
          2780..2801
          /rpt_family="AT_rich"
        repeat_region
          2836..2978
          /rpt_family="B1_MM"
        repeat_region
          complement(3017..3088)
          /rpt_family="ID_B1"
        repeat_region
          complement(3145..3259)
          /rpt_family="PBID9"
        repeat_region
          complement(3277..3388)
          /rpt_family="ID_B1"
        repeat_region
          3501..3725
          /rpt_family="Lx4"
        repeat_region
          3761..3799
          /rpt_family="AT_rich"
        repeat_region
          complement(3883..4054)
          /rpt_family="B2_Mm2"
        repeat_region
          4171..4315
          /rpt_family="B1_MM"
        repeat_region
          4568..4685
          /rpt_family="RSINE1"
        repeat_region
          4709..4797
          /rpt_family="ID_B1"
        repeat_region
          complement(5237..5358)
          /rpt_family="ID_B1"
        repeat_region
          5359..5386
          /rpt_family="(ATTG)n"
        repeat_region
          complement(5388..5526)
          /rpt_family="B2_Mm2"
        repeat_region
          5606..5806
          /rpt_family="B3"
        repeat_region
          5883..5913
          /rpt_family="(TTTG)n"
        repeat_region
          complement(5931..6122)
          /rpt_family="B2_Mm2"
        repeat_region
          6126..6167
          /rpt_family="(T)n"
        repeat_region
          7189..7377
          /rpt_family="ID_B1"
        repeat_region
          7378..7439
          /rpt_family="(GGA)n"
        repeat_region
          7467..7578
          /rpt_family="B4A"
        repeat_region
          7637..7720
          /rpt_family="B4A"

```

```

repeat_region
  7937..8061
  /rpt_family="B1_MM"
  complement(8243..8351)
  /rpt_family="PBID9"
  9410..9495
  /rpt_family="ID3"
  9540..9565
  /rpt_family="(TG)n"
  9691..9955
  /rpt_family="B4"
  10820..10679
  /rpt_family="PBI"
  10906..11053
  /rpt_family="B1_MM"
  complement(10966..10970)
  /note="<30 qual SNGL region"
  complement(11300..11422)
  /rpt_family="PBID9"
  11456..11585
  /rpt_family="RSINE1"
  complement(11674..11849)
  /rpt_family="MER33"
  complement(12141..12277)
  /rpt_family="B1_MM"
  complement(12294..12474)
  /rpt_family="B2_Mm2"
  12865..12696
  /rpt_family="(TC)n"
  12696..12716
  /rpt_family="(TG)n"
  13278..13424
  /rpt_family="B1_MM"
  15216..15389
  /rpt_family="B3A"
  15385..15510
  /rpt_family="B1_MM"
  15539..15674
  /rpt_family="RSINE1"
  15675..15723
  /rpt_family="B2_Mm1"
  16114..16168
  /rpt_family="GA-rich"
  16217..16445
  /rpt_family="URRIA"
  16499..16647
  /rpt_family="B4"
  17042..17106
  /rpt_family="(TG)n"

```

Alignment Scores:

Pred. No.:	1e+04	Length:	245002
Score:	45.00	Matches:	8
Percent Similarity:	90.91%	Conservative:	2
Best Local Similarity:	72.73%	Mismatches:	1
Query Match:	56.96%	Indels:	0
DB:	10	Gaps:	0

US-09-737-297-3 (1-16) x AC113276 (1-245002)

QY 2 GlucySerThr***AspValTyrGlnAsnIle 12

DB 90049 GAAGCTCCACTATGGACATCTACCAATCTATT 90017

RESULT 30

AC113273/c

LOCUS

DEFINITION

AC113273

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AC113273 245312 bp DNA linear HTG 23-APR-2003
Mus musculus clone RP23-380B8, WORKING DRAFT SEQUENCE, 3 unordered pieces.

AC113273.2 GI:30018004
HTG; HTGS PHASE1; HTGS DRAFT.
Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 245312)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-380B8
Unpublished
REFERENCE
AUTHORS

1 (bases 1 to 245312)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArelano,K., Dewari,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Gande,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,
Muegan,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Ratta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (28-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 245312)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Bouckgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArelano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagoopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrum,J., Meneus,L., Mihova,T., Mlangwa,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 17, 2003 this sequence version replaced gi:18997533.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997).
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L22703
Center clone name: 380_B_8

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 77164: contig of 77164 bp in length
77165 77264: gap of 100 bp
77265 90376: contig of 13112 bp in length
90377 90477: gap of 100 bp
90477 245312: contig of 154836 bp in length.

FEATURES

Location/Qualifiers
1..245312
/organism="Mus musculus"
/db_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-380B8"
/clone_lib="RPCI-23 Female Mouse BAC"
1..77164
/note="assembly_fragment"
clone end:SP6
vector_side:left
77265..90376
/note="assembly_fragment"
90477..245312
/note="assembly_fragment"

misc_feature

misc_feature

misc_feature

ORIGIN

Alignment Scores:
Pred. No.: 1e+04 Length: 245312
Score: 45.00 Matches: 8
Percent Similarity: 90.91% Conservative: 2
Best Local Similarity: 72.73% Mismatches: 1
Query Match: 56.96% Indels: 0
DB: 2 Gaps: 0

US-09-737-297-3 (1-16) x AC113273 (1-245312)

QY 2 GluglySerThr***AspValTyrGlnAsnIle 12

Db 103272 GAAGGCTCCATGACATGACATACCAATCTATT 103240

RESULT 31

AC113722/c

LOCUS

DEFINITION

AC113722

AC113722.5

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

Hollins, B., Howells, S., Hulyk, S., Hume, J., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, S., Jiang, H., Johnson, B., Johnson, R., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowals, C., Kratt, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, X., London, P., Longacre, S., Lopez, J., Lorensu, L., Loulseghe, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milsavljivic, A., Miner, G., Minja, E., Montanavor, J., Moore, S., Morgan, W., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okunolu, G., Olarunpase, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quintero, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, J., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, J., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1 to 251495)

Worley, K.C.

Direct Submission

Submitted (05-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 251495)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 19, 2002 this sequence version replaced gi:23267219.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Genome Center

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GSCT

Center clone name: CH230-63B2

Assembly program: Phrap, version 0.990329

Consensus quality: 235480 bases at least Q40

Consensus quality: 242469 bases at least Q30

Consensus quality: 244336 bases at least Q20

Estimated insert size: 246585; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 125238: contig of 125238 bp in length
* 125239 125338: gap of unknown length
* 125339 135027: contig of 9688 bp in length
* 135027 135126: gap of unknown length
* 135127 237172: contig of 102046 bp in length
* 237173 237272: gap of unknown length
* 237273 251495: contig of 14223 bp in length.

FEATURES

source

1..251495

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clones="CH230-63B2"

misc_feature

1..1380

/note="wgs end extension"

clone_end:T7

535..7632

/note="clone boundary"

clone_end:T7

site:

end sequence: BH315162"

1882..3705

/note="wgs_contig"

ORIGIN

Alignment Scores:
Pred. No.: 1.03e+04 Length: 251495
Score: 45.00 Matches: 8
Percent Similarity: 73.33% Conservative: 3
Best Local Similarity: 53.33% Mismatches: 4
Query Match: 56.96% Indels: 0
DB: 2 Gaps: 0

US-09-737-297-3 (1-16) x AC113722 (1-251495)

QY 2 GluGlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16

DB 229301 GAAGGATCTGTATGATGATTCCTCAAGTATATACAGCAAGGG 229257

RESULT 32

AC109570/c

LOCUS

DEFINITION

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

AC109570

Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, T., Foster, P., Fraser, C.W., Gabisi, A., Ganta, R., Garcia, A., Garner, M., Garza, M., Gregoridis, B., Geier, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A., Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowals, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewar, L., Louisedge, H., Lozada, R.J., Lu, X., Ma, J., Maneshwar, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawliny, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakoelameh, O., Okwuon, G., Olarunpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plapper, F., Poindexter, A., Popovic, D., Primus, E., Fu, L.-L., Puzos, W., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajis, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabors, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished 2 (bases 1 to 259329)
Worley, K.C.

Submitted (05-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Rat Genome Sequencing Consortium.

Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 19, 2002 this sequence version replaced gi:23267722. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information

Center project name: GQVH

Center clone name: CH230-53016
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 244964 bases at least Q40
Consensus quality: 247487 bases at least Q30
Consensus quality: 249131 bases at least Q20
Estimated insert size: 253869; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 7696: contig of 7696 bp in length
* 7697: gap of unknown length
* 7797: contig of 9701 bp in length
* 17498: gap of unknown length
* 17598: contig of 217029 bp in length
* 234626: gap of unknown length
* 234627: gap of unknown length
* 234727: contig of 22663 bp in length
* 257389: gap of unknown length
* 257390: contig of 1840 bp in length.
* 257490: contig of 1840 bp in length.

FEATURES
Location/Qualifiers
1..259329
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-53016"
complement(7726..8162)
/note="clone_boundary
clone_end:77
site:
end_sequence: BH344394"
7797..9695
/note="wgs_contig"

misc_feature
misc_feature
ORIGIN
Alignment Scores:
Pred. No.: 1.07e+04 Length: 259329
Score: 45.00 Matches: 9
Percent Similarity: 76.32% Conservative: 1
Best Local Similarity: 69.23% Mismatches: 3
Query Match: 56.96% Indels: 0
DB: 2 Gaps: 0
US-09-737-297-3 (1-16) x AC109570 (1-259329)
QY 4 SerThr***AppValTyrglnAsnIleGlnTyraAGly 16
Db 73840 AGCACTACTGATACATTCACAAACATCCATATCTCGG 73802

RESULT 33
AC117305
LOCUS
DEFINITION
Rattus norvegicus clone CH230-207C20, *** SEQUENCING IN PROGRESS
*** 6 unordered pieces.
ACCESSION
AC117305
VERSION
AC117305.4 GI:23194540
KEYWORDS
HTG, HTGS PHASE3, HTGS DRAFT, HTGS_ENRICHED.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1. (bases 1 to 302214)
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,

shotgun sequence only contigs will be indicated in the feature table.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GTQJ
Center clone name: CH230-207C20

Assembly program: Phrap; version 0.990329
Consensus quality: 188505 bases at least Q40
Consensus quality: 193334 bases at least Q30
Consensus quality: 196882 bases at least Q20
Estimated insert size: 192417; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

23373: contig of 22373 bp in length
22473: Gap of unknown length
22474: 292219: contig of 289746 bp in length
292220 292319: Gap of unknown length
292320 293801: contig of 1482 bp in length
293802 293901: Gap of unknown length
293902 295328: contig of 1427 bp in length
295329 295428: Gap of unknown length
295429 296692: contig of 1264 bp in length
296693 296792: Gap of unknown length
296793 302214: contig of 5422 bp in length.

FEATURES

source
1..302214
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-207C20"
misc_feature
1..1063
/note="wgs_end extension
clone_end:Sp6"
misc_feature
8244..9143
/note="clone boundary
clone_end:Sp6
site,EcoRI
end_sequence:RWBKJ22TVC"
misc_feature
1512..16383
/note="wgs_contig"
22474..24521
misc_feature
/note="wgs_contig"
complement(247891..248810)
/note="clone boundary
clone_end:77
site,EcoRI
end_sequence:RWBKJ22TJC"

ORIGIN

Alignment Scores:
Pred. No.: 1.27e+04 Length: 302214
Score: 45.00 Matches: 8
Percent Similarity: 91.67% Conservative: 3
Best Local Similarity: 66.67% Mismatches: 1
Query Match: 56.96% Indels: 0
DB: 2 Gaps: 0

US-09-737-297-3 (1-16) x AC117305 (1-302214)

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Avallebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, C., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z.,
Chuv, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, C., Denson, S., Derrico, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flegg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabris, A., Gant, R., Garcia, A., Garner, I., Garza, M.,
Gubaregova, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hayes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idelbird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpachy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebowitz, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenz, H., L., Louis, G., Lozano, R., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Plankoch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojase, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajda, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausen, A., Weis, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.

TITLE
JOURNAL
REFERENCE

2 (bases 1 to 302214)

Unpublished

Worley, K.C.

Direct Submission

Submitted (10-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 302214)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (10-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Sep 19, 2002 this sequence version replaced gi:21745821.

The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome

KEYWORDS mitochondrion Perca fluviatilis (perch)

SOURCE Perca fluviatilis

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae; Percidae; Perca.

REFERENCE 1 (bases 1 to 947)

AUTHORS Moyer, G.R., Sloss, B.L., Krajewski, C. and Billington, N.

TITLE Phylogenetic relationships among Perca species (Teleostei: Percidae) inferred from mtDNA sequence data

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 947)

AUTHORS Moyer, G.R., Sloss, B.L., Krajewski, C. and Billington, N.

TITLE Direct Submission

JOURNAL Submitted (12-SEP-2002) Zoology, Southern Illinois University, Life Science II, Carbondale, IL 62901, USA

FEATURES

source location/Qualifiers

1..947

/organism="Perca fluviatilis"

/organelles="mitochondrion"

/mol_type="genomic DNA"

/db_xref="taxon:8168"

/haplotype="2"

1..>947

/product="small subunit ribosomal RNA"

ORIGIN

Alignment Scores:

	30..9	Length:	947
Score:	44.00	Matches:	9
Percent Similarity:	60.00%	Conservative:	0
Best Local Similarity:	60.00%	Mismatches:	6
Query Match:	55.70%	Indels:	0
DB:	5	Gaps:	0

US-09-737-297-3 (1-16) x AF546105 (1-947)

QY 2 GluGlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16

DB 687 GAACGAAACATTAGATACCTACATAAACATTCAATACCTGGG 731

RESULT 36

CEW01P3/C

LOCUS CEW01P3

DEFINITION Caenorhabditis elegans cosmid W01P3, complete sequence.

ACCESSION Z92815

VERSION Z92815.1

KEYWORDS HTG.

SOURCE GI:2814509

ORGANISM Caenorhabditis elegans

Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;

Rhabditidae; Rhabditidae; Pelodidae; Caenorhabditis.

REFERENCE 1

AUTHORS none.

TITLE Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium

JOURNAL Science 282 (5396), 2012-2018 (1998)

99069613

REMARK The C. elegans Sequencing Consortium.

REFERENCE 2 (bases 1 to 32797)

AUTHORS Cummings, P.N.

TITLE Direct Submission

JOURNAL Submitted (09-MAR-1997) Nematode Sequencing Project, Sanger Institute, Hinxton, Cambridge CB10 1SA, England and Department of Genetics, Washington University, St. Louis, MO 63101, USA. E-mail: jesse@sanger.ac.uk or rwnematode.wustl.edu

ON Jan 28, 1998 this sequence version replaced gi:2414322.

Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.

COMMENT Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced

unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.

For a graphical representation of this sequence and its analysis see: <http://wormbase.sanger.ac.uk/per1/ace/elegans/seq/sequence?name=W01F3>

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

This sequence is the entire insert of clone W01F3. The true left end of clone F5398 is at 18053 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence AJ021491. The end of this sequence (32690..32797) overlaps with the start of sequence Z81547.

FEATURES

1..32797
 location/Qualifiers
 /organism="Caenorhabditis elegans"
 /mol_type="genomic DNA"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /chromosome="v"
 /clone="W01F3"
 Join(5411..5462,6547..6681,6890..7051,7660..7805,7854..8144,8645..8764,8812..9114,9321..9438,9721..9832,10087..10240)
 /gene="W01F3.1a"
 Join(5411..5462,6547..6681,7660..7805,7854..8144,8645..8764,8812..9114,9321..9438,9721..9832,10087..10240)
 /gene="W01F3.1b"
 Join(5411..5462,6547..6681,6890..7051,7660..7805,7854..8144,8645..8764,8812..9114,9321..9438,9721..9832,10087..10240)
 /gene="W01F3.1a"
 /standard_name="W01F3.1a"
 /note="cDNA EST cml3a3 comes from this gene"
 cDNA EST YK291a9.5 comes from this gene
 cDNA EST YK281b11.5 comes from this gene
 cDNA EST YK394c6.5 comes from this gene
 cDNA EST YK6229f1.5 comes from this gene
 cDNA EST YK509b6.5 comes from this gene"
 /codon_start=1
 /product="Hypothetical protein W01F3.1a"
 /protein_id="CAB07292.2"
 /db_xref="GI:13548410"
 /translation="MMISILLVFAVSTAYCRDYGRENRAIRMSVDFWKNKTKDGLNVRVAGVNSWPDLDVQLHEPFLNKMIEGVFMQARRRPRHRAVPVFNKADTKLADRI TFKEVREVTAKAIKINFESNEVRVPLQDYASQADTFVLPQYVILHANFRS PRSVTEYDPTFVEKADLPDPTATLSLGTQKSNFSLHFKFKLSWQLFHILEYI SRLDQVMSVRLSVAAHSKQLLWLMQDSISLLWSADHDVNTWTPIVELRST TKNRLYDLDPKHKRFLQTDNEPIATPTFSLQDQWQAVFSSGSLSTVVRSEKGP AFLGPEALLSKTPPPNFPSEQKISGKHVFLPKKGMENVDLDSGVSILWMDKVQD IDSPKINSLSVEIFGFGKIKITNGQLKNLPYETKSYGQLPDSDCYGEVETMDGWRV KLDWATACGQKQKRNKRTVSIELDTPFQKRRHNRNVVSKGSDAVDFLLBELQHRN AYNRNPTVTVSTILLMIYLL"
 Join(5411..5462,6547..6681,7660..7805,7854..8144,8645..8764,8812..9114,9321..9438,9721..9832,10087..10240)
 /gene="W01F3.1b"
 /standard_name="W01F3.1b"
 /note="cDNA EST YK291a9.3 comes from this gene"
 cDNA EST YK394c6.3 comes from this gene
 cDNA EST YK458e9.3 comes from this gene
 cDNA EST YK458e9.5 comes from this gene
 cDNA EST YK6229f1.3 comes from this gene
 cDNA EST YK509b6.3 comes from this gene"
 /codon_start=1
 /product="Hypothetical protein W01F3.1b"
 /protein_id="CAB07293.2"
 /db_xref="GI:13548411"
 /translation="MMISILLVFAVSTAYCRDYGRENRAIRMSVDFWKNKTKDGLNVRVAGVNSWPDLDVQLHAIKINRNEVRVPLQDYASQADTFVLPQYVILHAN FRSPRSVTEYDPTFVEKADLPDPTATLSLGTQKSNFSLHFKFKLSWQLFHILEYI LYSRLDQVMSVRLSVAAHSKQLLWLMQDSISLLWSADHDVNTWTPIVEL

RRSTKNRIYLDLPKHKRFLQTDNEPIATPTFSLQDQWQAVFSSGSLSTVVRSEKPAFLGPTALLSKIPPPNFPSEQKISGKHVFLPKKGMENVDLDSGVSILWMDKVQDIDSPFKINSLEVFIFGDKITNGQLKNLPYETKSYGQLPDSDCYGEVETMDGWRVLDWATACGQKQKRNKRTVSIELDTPFQKRRHNRNVVSKGSDAVDFLLBELQHRNAYNRPVTVSTILLMIYLL"
 Join(11114..11200,11241..11476,11666..12038,12381..12458,12505..12654)
 /gene="W01F3.2"
 Join(11114..11200,11241..11476,11666..12038,12381..12458,12505..12654)
 /gene="W01F3.2"
 /standard_name="W01F3.2"
 /note="cDNA EST YK163c6.5 comes from this gene"
 cDNA EST YK163c6.3 comes from this gene
 cDNA EST YK234d4.3 comes from this gene
 cDNA EST YK252h1.3 comes from this gene
 cDNA EST YK367c12.3 comes from this gene
 cDNA EST YK444g5.3 comes from this gene
 cDNA EST YK234d4.5 comes from this gene
 cDNA EST YK252h1.5 comes from this gene
 cDNA EST YK444g5.5 comes from this gene"
 /codon_start=1
 /product="Hypothetical protein W01F3.2"
 /protein_id="CAB07291.1"
 /db_xref="GI:3880423"
 /translation="MRLSILLVAHVANGLFGFGKKTTOAPKISLGSEISIPHTDE DQGTGEIVIPENDPSTTTRRTATTGTGIPVSWGGIGLDRTESVTVDRSFTTLNPR FTAATKRLPSDDHSSSGCPNRIDATFPGDSDYAEYEQVKISNRNSDRTILVEE FKQPNVGNALYDPEREILLWIDGRSVYIGKDGSDNNKLQSVFPKELPSSIGFTPE AAVRNHNKHSLLSSNGKFPALYDEYWNKSLMTGRTESYFENPDRVGRGISTWNSQHA HIYQSLVFFYDSEMKKVTGQDGVPLGDFWRC"
 Join(15973..16069,17927..17995,18045..18176,18231..18358,18432..18818,18889..19426,19476..19699,19743..20487,20723..21679,21733..21987,25147..25275,25321..25605,25657..27360,27407..27631,27680..28294,28719..28872,28985..29018)
 /gene="W01F3.3"
 Join(15973..16069,17927..17995,18045..18176,18231..18358,18432..18818,18889..19426,19476..19699,19743..20487,20723..21679,21733..21987,25147..25275,25321..25605,25657..27360,27407..27631,27680..28294,28719..28872,28985..29018)
 /gene="W01F3.3"
 /standard_name="W01F3.3"
 /note="contains similarity to Pfam domain: PF0001.4 (Kunitz/Bovine pancreatic trypsin inhibitor domain), Score=696.1, E-value=5.9e-226, N=1; PF00086 (Thyroglobulin type-1 repeat), Score=72.8, E-value=2.3e-18, N=1"
 cDNA EST CEES192f comes from this gene; cDNA EST YK110d9.3 comes from this gene
 cDNA EST YK110d9.5 comes from this gene; cDNA EST YK53f7.5 comes from this gene
 cDNA EST YK188a8.3 comes from this gene; cDNA EST YK188a8.5 comes from this gene
 cDNA EST YK248c12.5 comes from this gene; cDNA EST YK269f5.5 comes from this gene
 cDNA EST YK313h10.5 comes from this gene; cDNA EST YK373a7.5 comes from this gene
 cDNA EST YK404a9.5 comes from this gene; cDNA EST YK442e4.5 comes from this gene
 cDNA EST YK445f8.5 comes from this gene; cDNA EST YK475g12.5 comes from this gene
 cDNA EST YK210g7.3 comes from this gene; cDNA EST YK238d8.3 comes from this gene
 cDNA EST YK260f2.3 comes from this gene; cDNA EST YK250a7.3 comes from this gene
 cDNA EST YK286h9.3 comes from this gene; cDNA EST YK290b2.3 comes from this gene
 cDNA EST YK312d3.3 comes from this gene; cDNA EST YK391f7.3 comes from this gene
 cDNA EST YK402d8.3 comes from this gene; cDNA EST YK448g2.3 comes from this gene

CDNA EST yk456h8.3 comes from this gene; CDNA EST yk431g10.3 comes from this gene
 CDNA EST yk495h12.3 comes from this gene; CDNA EST yk497g9.3 comes from this gene
 CDNA EST yk501e10.3 comes from this gene; CDNA EST yk210g7.5 comes from this gene
 CDNA EST yk238d8.5 comes from this gene; CDNA EST yk260f2.5 comes from this gene
 CDNA EST yk250a7.5 comes from this gene; CDNA EST yk286h9.5 comes from this gene
 CDNA EST yk290b2.5 comes from this gene; CDNA EST yk312d3.5 comes from this gene
 CDNA EST yk391f7.5 comes from this gene; CDNA EST yk402d8.5 comes from this gene
 CDNA EST yk448g2.5 comes from this gene; CDNA EST yk456h8.5 comes from this gene
 CDNA EST yk495h12.5 comes from this gene; CDNA EST yk497g9.5 comes from this gene
 CDNA EST yk501e10.5 comes from this gene; CDNA EST yk502g7.5 comes from this gene

Alignment Scores:

Pred. NO.: 1.66e+03 Length: 32797
 Score: 44.00 Matches: 10
 Percent Similarity: 81.25% Conservative: 3
 Best Local Similarity: 62.50% Mismatches: 1
 Query Match: 55.70% Indels: 2
 DB: 3 Gaps: 1

US-09-737-297-3 (1-16) x CEW01P3 (1-32797)

Qy 2 GlucySerThr***AspValTyGlnAsnIle-----GlnTyTala 15

DB 26947 GARGGCACATCACTGGACGTCCTACGATATCTGACTTCAGTACGCC 26800

RESULT 37

HSTCRBA120/c HSTCRBA120 36059 bp DNA linear PRI 24-JUL-1997
 LOCUS HSTCRBA120 36059 bp DNA linear PRI 24-JUL-1997
 DEFINITION Homo sapiens T cell receptor beta locus, TCRBV6S4A1 to TCRBV8S1

ACCESSION AF009661
 VERSION AF009661.1 GI:2275567

KEYWORDS
 SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 36059)
 Rowen, L., Sato, J., Smit, A., Acharya, C., Ahearn, M.E., Ankener, M., Baskin, D., Bumgarner, R., Chen, L., Chen, N., Deshpande, P., Faust, J., Howard, S., Jerome, N., Koop, B.F., Lee, H., Loretz, C., Paepker, B., Zackrone, K. and Hood, L.

Sequence determination of the human T cell receptor beta locus: Strategy and error analysis

JOURNAL

Unpublished

2 (bases 1 to 36059)

Rowen, L., Wang, X., Boysen, C., Ahearn, M.E., Charnley, P., Paepker, B.,

Lee, I., Chen, L., Trask, B., Nickerson, D., Sato, D. and Hood, L.

Sequence variation among several haplotypes in the T cell receptor

JOURNAL

Unpublished

3 (bases 1 to 36059)

Rowen, L.

Direct Submission

Submitted (23-JUN-1997) Department of Molecular Biotechnology, Box

357730 University of Washington, Seattle, Washington 98195, USA

This contig overlaps positions 360292-396358 in the original TCRB

entry found in GSD Accession Number L36092 and positions

102774-138840 in Genbank Accession Number U66060. Sequencing methodology: high redundancy shotgun. Interspersed Repeats were identified with RepeatMasker (available from <http://ftp.genome.washington.edu/RM/RepeatMasker.html>) Simple sequence repeats were identified with sputnik (available from <http://setac.mbt.washington.edu/~chrisa/software/sputnik.html>).

FEATURES

source

1. 36059
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

/map="7q35"

/note="cosmid A1-20, from YAC 234 A6F6, CGM1 cell line, haplotype A"

1. 32021

/note="Overlap with cosmid K34, YAC D49H4, CGM1 cell line, haplotype B"

467

/note="cosmid A1-20: a; cosmid K34: g"

/replace="g"

796

/note="cosmid A1-20: t; cosmid K34: c"

/replace="c"

903

/note="cosmid A1-20: a; cosmid K34: g"

/replace="g"

2140

/note="cosmid A1-20: c; cosmid K34: t"

/replace="t"

3733. >4357

/gene="TCRBV6S4A1"

/note="proposed new name: TCRBV7S9"

3733. 3748

/gene="TCRBV6S4A1"

/note="conserved 16mer; possible promoter"

Join(3846..3894,4021..4318)

/gene="TCRBV6S4A1"

/codon_start=1

/product="TCRBV6S4"

/protein_id="AAB63895.1"

/db_xref="GI:2275568"

/translation="MTSLTCLWALCLLGADHDTGVSONPRHKITKRGQNVTFRCDDP

ISEHRLYWRQTLQGQPEFLTYFQNAQLEKSRLLSDRFSAERPKGSPSTLQIQRTE

QGDGAMYLCASSLA"

Join(3846..3894,4021..4318)

/gene="TCRBV6S4A1"

/standard_name="TCRBV6S4"

3940..3974

/note="microsatellite"

/rpt_type=tandem

/rpt_unit="TG"

4047

/gene="TCRBV6S4A1"

/note="cosmid A1-20: a; cosmid K34: g"

/replace="g"

4061..36059

/note="Overlap with cosmid H7.1, isolated from sperm"

4319..4325

/gene="TCRBV6S4A1"

/note="heptamer"

4326..4348

/gene="TCRBV6S4A1"

/note="spacer"

4349..4357

/gene="TCRBV6S4A1"

/note="nonamer"

4634

/note="cosmid A1-20: t; cosmids K34 and H7.1: c"

/replace="c"

complement(5062..5118)

/rpt_family="LIME2"

5126

/note="cosmid A1-20: t; cosmids K34 and H7.1: g"

repeat_region

variation

misc_feature

misc_recomb

misc_recomb

misc_recomb

variation

repeat_region

variation


```
repeat_region /replace="g"
variation      complement(5134..5709)
              /rpt_family="L1"
5546          /note="cosmids A1-20 and K34: t; cosmid H7.1: c"
              /replace="c"
5640          /note="cosmid A1-20: c; cosmids K34 and H7.1: t"
              /replace="t"
5747          /note="cosmid A1-20: t; cosmid K34 and H7.1: tt"
              /replace="tt"
5764..6267    /rpt_family="L191_5end"
5917          /note="cosmid A1-20: g; cosmids K34 and H7.1: t"
              /replace="t"
5921          /note="cosmid A1-20: a; cosmids K34 and H7.1: g"
              /replace="g"
6052          /note="cosmid A1-20: a; cosmids K34 and H7.1: g"
              /replace="g"
6114          /note="cosmid A1-20: c; cosmids K34 and H7.1: g"
              /replace="g"
6200          /note="cosmids A1-20 and K34: g; cosmid H7.1: gg"
              /replace="gg"
6278..6282    /note="cosmid A1-20: tgaat; cosmids K34 and H7.1: t"
              /replace="t"
6330          /note="cosmid A1-20: a; cosmids K34 and H7.1: c"
              /replace="c"
6482          /note="cosmid A1-20: a; cosmids K34 and H7.1: at"
              /replace="at"
6554          /note="cosmid A1-20: a; cosmids K34 and H7.1: t"
              /replace="t"
6798..7351    /rpt_family="L1"
7221          /note="cosmid A1-20: g; cosmids K34 and H7.1: a"
              /replace="a"
7230          /note="cosmid A1-20: g; cosmids K34 and H7.1: t"
              /replace="t"
7344          /note="cosmid A1-20: g; cosmids K34 and H7.1: a"
              /replace="a"
7541          /note="cosmid A1-20: t; cosmids K34 and H7.1: tt"
              /replace="tt"
7553          /note="cosmid A1-20: g; cosmids K34 and H7.1: c"
              /replace="c"
7604          /note="cosmid A1-20: c; cosmids K34 and H7.1: t"
              /replace="t"
7661          /note="cosmid A1-20: t; cosmids K34 and H7.1: c"
              /replace="c"
7744          /note="cosmid A1-20: c; cosmids K34 and H7.1: g"
              /replace="g"
7853          /note="cosmid A1-20: g; cosmids K34 and H7.1: t"
              /replace="t"
8092..8446    repeat_region
```

```
variation      /rpt_family="L1"
8397          /note="cosmid A1-20: g; cosmids K34 and H7.1: a"
              /replace="a"
8448..8449    /note="cosmid A1-20: aa; cosmids K34 and H7.1: a"
              /replace="a"
8573          /note="cosmid A1-20: a; cosmids K34 and H7.1: g"
              /replace="g"
8621..8622    /note="cosmids A1-20 and K34: aa; cosmid H7.1: a"
              /replace="a"
complement(8677..9001)
9030          /rpt_family="MER58B"
          /note="cosmid A1-20: a; cosmids K34 and H7.1: aa"
              /replace="aa"
9072          /note="cosmid A1-20: c; cosmids K34 and H7.1: cc"
              /replace="cc"
9383          /note="cosmids A1-20 and K34: c; cosmid H7.1: ac"
```

```
Alignment Scores:
Pred. No.:      1.95e+03      Length:      36059
Score:          44.00         Matches:    10
Percent Similarity: 73.33%    Conservative: 1
Best Local Similarity: 66.67%  Mismatches:  4
Query Match:      55.70%     Indels:      0
DB:               9          Gaps:         0
```

US-09-737-297-3 (1-16) x HSTCRBA120 (1-36059)

```
QY  2  GluGlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16
      |||||  ::|||  |||||  |||||  |||||  |||||  |||||  |||||
Db   17979  GAAAGATCTACTCTCAATGTGTATGGCAATATCCAGTCAGTCGGG 17935
```

```
RESULT 38
AL445187  76114 bp  DNA  linear  PRI 05-DEC-2001
LOCUS    Human DNA sequence from clone RP11-576C12 on chromosome 9 Contains
DEFINITION a VPS33B pseudogene, complete sequence.
ACCESSION AL445187
VERSION   AL445187.7  GI:11544549
KEYWORDS  HTG.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Corby,N.
```

```
REFERENCE 1
AUTHORS   Direct Submission
TITLE     Submitted (05-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL   Cambridgehire, CB10 1SA, UK. E-mail enquiries:
          humquy@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
          On Dec 4, 2000 this sequence version replaced GI:11340318.
          During sequence assembly data is compared from overlapping clones.
          Where differences are found these are annotated as variations.
          A variation annotation may not be found in the sequence submission
          corresponding to the overlapping clone, as we submit sequences with
          only a small overlap as described above.
          The following abbreviations are used to associate primary accession
          numbers given in the feature table with their source databases:
          Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information
          on the WORMPEP database can be found at
          http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
          was generated from part of Bacterial clone contigs of human
          chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
          Group. Further information can be found at
          http://www.sanger.ac.uk/HGP/Chr9
          This sequence was finished as follows unless otherwise noted: all
```



```

/notes="Alusx repeat: matches 1. .301 of consensus"
complement(31945. .31963)
/notes="Sequence from overlapping clone ba276E15
(AL1390067) Assembly confirmed by restriction digest."
31995. .32288
/notes="Alusp repeat: matches 1. .295 of consensus"
33495. .33681
/notes="MIR repeat: matches 9. .205 of consensus"
complement(33899. .34718)
/notes="match: GSS: Em: AQ743239"
33999. .34064
/notes="33 copies 2 mer tt 66% conserved"
34080. .34720
/notes="L2 repeat: matches 2080. .2710 of consensus"
34993. .35270
/notes="AluJo repeat: matches 1. .274 of consensus"
complement(36692. .37430)
/notes="match: GSS: Em: AQ481167"
37056. .37286
/notes="MIR repeat: matches 1. .232 of consensus"
37287. .37427

Alignment Scores:
Pred. No.: 4.28e+03 Length: 76114
Score: 44.00 Matches: 7
Percent Similarity: 78.57% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 3
Query Match: 55.70% Indels: 0
DB: 9 Gaps: 0

US-09-737-297-3 (1-16) x AL445187 (1-76114)

Qy 2 GlucylserThr***AspValtyrGlnAsnileGlnTyrAla 15
:::|||||:::
Db 46897 GATGGTCCACAGTGATCTTACCAGATTTCAGTATAGT 46938

RESULT 39
HSTCRBV/c HSTCRBV 77743 bp DNA linear PRI 20-APR-1994
LOCUS Human v beta T-cell receptor (TCRBV) gene locus.
DEFINITION U03115
ACCESSION U03115
VERSION U03115.1 GI:467918
KEYWORDS Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 77743)
Slightom,J.L., Siemieniak,D.R., Sieu,L.C., Koop,B.F. and Hood,L.
Nucleotide sequence analysis of 77.7 kb of the human v beta T-cell
receptor gene locus: direct primer-walking using cosmid template
DNAs
Genomics 20 (2), 149-168 (1994)
JOURNAL
MEDLINE 94292194
PubMed 8020962
REFERENCE 2 (bases 1 to 77743)
Slightom,J.L.
AUTHORS Direct Submission
TITLE Submitted (04-NOV-1993) Jerry L. Slightom, Molecular Biology Unit,
The Upjohn Company, 301 Henrietta Street, Kalamazoo, MI 49007, USA
FEATURES
source
1. 77743
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/strain="adult cell cultures or sperm fibroblast"
/db_xref="taxon:9606"
/clone="cosmid clones H7.1, H12.18, and H130.1"
1. .253
/gene="TCRVB6S1"
<i>..>253
/gene="TCRVB6S1"
/codon_start=2
/product="T-cell receptor beta chain V region precursor"

```

```

/protein_id="AAAI7709.1"
/db_xref="GI:467919"
/translation="ITKGGQNVTERCDPISEHNRLYWYRQTLGQGPFLTYFQNEAQL
EKSRLLSDRPSAERPKGSPFSLFIQRTEQDSAMYLCASS"
<i>..>253
/gene="TCRVB6S1"
/number=2
complement(3484. .3772)
/rpt_family="Alu Class Sx"
6305. .6783
/gene="TCRBV23S1"
join(6305. .6383,6494. .>6783)
/gene="TCRBV23S1"
/codon_start=1
/product="T-cell receptor beta chain V region precursor"
/protein_id="AAAI7710.1"
/db_xref="GI:467920"
/translation="MLSPDLFPDSAWNTRLLCHVLMCLLGAVSVAAGVIOSPRHILKEK
RETATLKCYPIPRHDTVYVYQGGQDPQLISFYEKMQSDKGSIPDRFSAQQPFSYH
SELNMSLEIGDSALYFCASS"
6305. .6383
/gene="TCRVB23S1"
6305. .6383
/gene="TCRVB23S1"
/number=1
6384. .6393
/gene="TCRBV23S1"
6494. .6783
/gene="TCRBV23S1"
/number=2
complement(7501. .7650)
/rpt_family="Alu Class j"
complement(8654. .8942)
/rpt_family="Alu Class sb"
11357. .11686
/rpt_family="Alu Class Sx"
14717. .15161
/gene="TCRBV12S2"
join(14717. .14765,14872. .>15161)
/product="T-cell receptor beta chain V region precursor"
/protein_id="AAAI7711.1"
/db_xref="GI:467921"
/translation="WGTRLPFYVALCLLWTHMDAGITQSPRHKVETGTPTVLRCHQ
TENHYMYWVRQDPGHGLRLIHYSYGVKDKDKGEVSDGYSVRSKTDFTLLLESATS
SQTSYVFCAS"
14717. .14765
/gene="TCRBV12S2"
14766. .14871
/number=1
/gene="TCRBV12S2"
/number=1
14872. .15161
/gene="TCRBV12S2"
/number=2
16004. .16293
/rpt_family="Alu Class Sq"
16582. .16867
/rpt_family="Alu Class Sc"
17843. .18135
/rpt_family="Alu Class Sx"
complement(18318. .18677)
/rpt_family="OFR"
complement(20097. .20939)
/rpt_family="Kpn LINE"
25365. .25797
/gene="TCRBV21S2"
join(25365. .25413,25505. .>25797)
/gene="TCRBV21S2"
/codon_start=1
/product="T-cell receptor beta chain V region precursor"
/protein_id="AAAI7712.1"

```


Percent Similarity: 90.91% Conservative: 2
 Best Local Similarity: 72.73% Mismatches: 1
 Query Match: 55.70% Indels: 0
 B: 2 Gaps: 0

IS-09-737-297-3 (1-16) x AC112592.1 (1-110000)
 Y 3 GlySerThr***AspValTyrGlnAsnIleGln 13
 b 26500 GGTTCAACTATTGACTTATATCAAAATATAAG 26532

RESULT 42
 AC067868 136434 bp DNA linear HTG 24-AUG-2002
 DEFINITION Homo sapiens chromosome 9 clone RP11-109119 map 9, WORKING DRAFT
 AC067868 SEQUENCE, 15 unordered pieces.
 AC067868 HTG: HTGS PHASE1: HTGS_DRAFT.
 JERSON
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 136434)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Homo sapiens chromosome 9, clone RP11-109119
 JOURNAL
 REFERENCE
 2 (bases 1 to 136434)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Boukagalter, B., Brown, A., Burkett, G.,
 Campiano, A., Castelle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collamore, A., Cooke, E., DeArrellano, K., Dewar, K., Diaz, J. S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lechowky, J.,
 Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, T., Norman, C. H., O'Connor, T., O'Donnell, P.,
 O'Neil, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 136434)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Boguslavsky, L., Boukagalter, B., Brown, A.,
 Camarata, J., Campiano, A., Chang, J., Choepel, Y., Colangelo, M.,
 Collins, S., Collamore, A., Cooke, P., DeArrellano, K., Dewar, K.,
 Diaz, J. S., Dodge, S., Fato, S., Ferreira, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W.,
 Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRoque, K.,
 Lamazares, R., Landers, T., Lechowky, J., Levine, R., Liu, G.,
 Maclean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M.,
 McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L.,
 Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C.,
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J.,
 Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
 Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S.,
 Severy, P., Sougniez, C., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S.,
 Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.

Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 16, 2000 this sequence version replaced gi:7651920.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WJBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center Project name: L8967
 Center Clone name: 109 I.19
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 126982 bases at least Q40
 Consensus quality: 132391 bases at least Q30
 Consensus quality: 134263 bases at least Q20
 Insert size: 133000; agarose-fp
 Insert size: 135034; sum-of-contigs
 Quality coverage: 4.4 in Q20 bases; agarose-fp
 Quality coverage: 4.3 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 15 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1544: contig of 1544 bp in length
 * 1545 1644: gap of 100 bp
 * 1645 3188: contig of 1544 bp in length
 * 3189 3288: gap of 100 bp
 * 3289 6688: contig of 3400 bp in length
 * 6689 6788: gap of 100 bp
 * 6789 9888: contig of 3100 bp in length
 * 9889 9988: gap of 100 bp
 * 9989 14056: contig of 4068 bp in length
 * 14057 14156: gap of 100 bp
 * 14157 19673: contig of 5517 bp in length
 * 19674 19773: gap of 100 bp
 * 19774 25908: contig of 6135 bp in length
 * 25909 26008: gap of 100 bp
 * 26009 29213: contig of 3205 bp in length
 * 29214 29313: gap of 100 bp
 * 29314 32293: contig of 6980 bp in length
 * 32294 36393: gap of 100 bp
 * 36394 45349: contig of 8956 bp in length
 * 45350 45449: gap of 100 bp
 * 45450 55061: contig of 9612 bp in length
 * 55062 55161: gap of 100 bp
 * 55162 71865: contig of 16704 bp in length
 * 71866 71965: gap of 100 bp
 * 71966 94253: contig of 22288 bp in length
 * 94254 94353: gap of 100 bp
 * 94354 115381: contig of 21028 bp in length
 * 115382 115481: gap of 100 bp
 * 115482 136434: contig of 20953 bp in length.
 Location/Qualifiers
 1. 136434
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="9"
 /map="9"
 /clone="RP11-109119"

FEATURES
source

misc_feature 1. 1544
 /clone lib="RPC1-11 Human Male BAC"
 /note="assembly_fragment"
 misc_feature 1845. 3188
 /note="assembly_fragment"
 misc_feature 3289. 6688
 /note="assembly_fragment"
 misc_feature 6789. 9888
 /note="assembly_fragment"
 misc_feature 9989. 14056
 /note="assembly_fragment"
 misc_feature 14157. 13673
 /note="assembly_fragment"
 misc_feature 19774. 25908
 /note="assembly_fragment"
 misc_feature 26009. 29213
 /note="assembly_fragment"
 /clone end:SP6
 vector side:left
 misc_feature 29314. 36293
 /note="assembly_fragment"
 misc_feature 36394. 45349
 /note="assembly_fragment"
 /clone end:T7
 vector side:right
 misc_feature 45450. 55061
 /note="assembly_fragment"
 misc_feature 55162. 71865
 /note="assembly_fragment"
 misc_feature 71866. 94253
 /note="assembly_fragment"
 misc_feature 94354. 115381
 /note="assembly_fragment"
 misc_feature 115482. 136434
 /note="assembly_fragment"

ORIGIN

Alignment Scores: 136434
 Pred. No.: 7
 Score: 8.26e+03
 Percent: 44.00
 Percent Similarity: 78.57%
 Best Local Similarity: 50.00%
 Query Match: 55.70%
 DB: 2
 Matches: 7
 Mismatches: 3
 Indels: 0
 Gaps: 0

US-09-737-297-3 (1-16) x AC067868 (1-136434)

QY 2 GluglySerThr***AspValTy-GlnAsnIleGlnTyAla 15

Db 70096 GATGTTCCACACGTGATTTCTACAGAAATTTCAAGTATAGT 70137

RESULT 43

HSTCRB75A/c HSTCRB75A 136975 bp DNA linear PRI 24-JUL-1997

LOCUS Homo sapiens T cell receptor beta locus, TCRBV21S2A2

DEFINITION

ACCESSION AF009663

VERSION AF009663.1 GI:2275578

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 136975)

Rowen, L., Seto, J., Smit, A., Acharya, C., Ahearn, M. B., Ankner, M.,

Baskin, D., Bumgarner, R., Chen, L., Chen, N., Deehpande, P., Faust, J.,

Howard, S., Jerome, N., Koop, B. F., Lee, H., Loretz, C., Paepers, B.,

Zackrohn, K., and Hood, L.

Sequence determination of the human T cell receptor beta locus:

Strategy and error analysis

Unpublished

REFERENCE 2 (bases 1 to 136975)

Rowen, L., Wang, K., Boyesen, C., Ahearn, M. B., Charnley, P., Paepers, B.,

Lee, I., Chen, L., Trask, B., Nickerson, D., Seto, D. and Hood, L.

Sequence variation among several haplotypes in the human T cell

receptor beta locus

Unpublished

Rowen, L.

Lee, I., Chen, L., Trask, B., Nickerson, D., Seto, D. and Hood, L.
 Sequence variation among several haplotypes in the human T cell
 receptor beta locus
 Unpublished
 3 (bases 1 to 136975)
 Rowen, L.
 Direct Submission
 Submitted (23-JUN-1997) Department of Molecular Biotechnology, Box
 357730 University of Washington, Seattle, Washington 98195, USA
 This contig was assembled from shotgun data from five overlapping
 cosmids (7-5A, 4-8, 2-9, CA33, K34) subcloned from YACs of the same
 haplotype of the CGM1 cell line. It overlaps the original TCRB
 entry found in GDB Accession Number L36092 from positions
 25531-39304. It overlaps Genbank Accession Number U66059 from
 positions 25531-262156 and U66060 from positions 1-134786. In
 this entry, only the sequence variations between CGM1 haplotypes B
 and A are annotated. For complete annotation of variations, see
 U66059 and U66060. Sequencing Methodology: high redundancy shotgun.
 Interspersed Repeats were identified with RepeatMasker (available
 from <http://ftp.genome.washington.edu/RM/RepeatMasker.html>) Simple
 sequence repeats were identified with sputnik (available from
<http://serac.mbt.washington.edu/~chrisa/software/sputnik.html>).

FEATURES

Location/Qualifiers
 1..136975
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /map="7Q35"
 /cell_line="CGM1"
 /note="Cosmid 7-5A, 4-8, and 2-9 are from YAC 196F4, CGM1
 haplotype B. Cosmids CA33 and K34 are from YAC D49H4,
 CGM1 haplotype B."
 1..38822
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /map="7Q35"
 /clone="cosmid 7-5A"
 /note="cosmid 7-5A is from YAC 196F4, CGM1 haplotype B"
 27445..64638
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /map="7Q35"
 /clone="cosmid 4-8"
 /note="cosmid 4-8 is from YAC 196F4, CGM1 haplotype B"
 43687..86549
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /map="7Q35"
 /clone="cosmid 2-9cosmid 2-9"
 /note="cosmid 2-9 is from YAC 196F4, CGM1 haplotype B"
 71498..104645
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /map="7Q35"
 /clone="cosmid CA33"
 /note="cosmid CA33 is from YAC D49H4, CGM1 haplotype B"
 96917..136975
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /map="7Q35"
 /clone="cosmid K34"
 /note="cosmid K34 is from YAC D49H4, CGM1 haplotype B"
 124..5722
 /gene="TCRBV85SP"
 /note="proposed new name: TCRBV12S2. This pseudogene
 contains an internal stop codon"
 /pseudo
 misc_feature 124..139


```

/genes="TCRBV6S5P"
/notes="conserved 16mer; possible promoter"
/pseudo
join(241..286,386..>683)
/genes="TCRBV8S5P"
/pseudo
/codon_start=1
join(241..286,386..683)
/genes="TCRBV8S5P"
/standard_name="TCRBV8S5"
/pseudo
684..690
/genes="TCRBV8S5P"
/notes="heptamer"
691..713
/genes="TCRBV8S5P"
/notes="spacer"
714..722
/genes="TCRBV8S5P"
/notes="nonamer"
1424..1652
/rpt_family="AluJo"
4222..40558
/notes="overlap with cosmid B27, CGM1 haplotype A"
5247..5481
/rpt_family="LIMA4"
complement(6221..6462)
/rpt_family="LIMB1"
6466..6762
/rpt_family="AluSq"
complement(6764..6899)
/rpt_family="LIMA10"
complement(7852..8058)
/rpt_family="LIME2"
8170..8581
/rpt_family="LIPA9"
10371..>10845
/genes="TCRBV13S1"
/notes="proposed new name: TCRBV6S5"
join(10371..10419,10512..>10806)
/genes="TCRBV13S1"
/codon_start=1
/product="TCRBV13S1"
/protein_id="AAB63900.1"
/db_xref="GI:2275579"
/translation="MSIGLLCCALSLWAGFVNAVGTQTPKFOVLKGTQSMILQCAQ
DMHMYMYRQDPGMLRLIHYSGAGITDQGEVPGVNGVNSRSTTDFPLRLISAAP
SOTSVPFCASSYS"
join(10371..10419,10512..10806)
/genes="TCRBV13S1"
/standard_name="TCRBV13S1"
10807..10813
/genes="TCRBV13S1"
/notes="heptamer"
10814..10836
/genes="TCRBV13S1"
/notes="spacer"
10837..10845
/genes="TCRBV13S1"
/notes="nonamer"
13590..13844
/rpt_family="L191_send"
14419..>15032
/genes="TCRBV6S8A2T"
/notes="proposed new name: TCRBV7S4"
14419..14434
/genes="TCRBV6S8A2T"
/notes="conserved 16mer; possible promoter"
join(14532..14580,14696..>14993)
/genes="TCRBV6S8A2T"
/product="TCRBV6S8"
/protein_id="AAB63903.1"

```

```

/db_xref="GI:2275582"
/translation="MGTRLLCQNVVLGFLGTDHTGAGVSQSPRYKVAKRGRDVALRCD
ISGSHLYMYRQITLQSGSEVLITYSQDAQRDKSGRPSGRFSABRPERSVSTLKIQRT
QGDSAVYLCASSLA"
join(14532..14580,14696..14993)
/genes="TCRBV6S8A2T"
/standard_name="TCRBV6S8"
14994..15000
/genes="TCRBV6S8A2T"
/notes="heptamer"
15001..15023
/genes="TCRBV6S8A2T"
/notes="spacer"
15024..15032
/genes="TCRBV6S8A2T"
/notes="nonamer"
complement(16818..17015)
/rpt_family="MER63A"
complement(20469..21005)
/rpt_family="LIME2"
21005..21366
/rpt_family="MSTD"
21400..21735
/rpt_family="MST-INTERNAL"
22357..>22372
/genes="TCRBV6S6A3N2T"
/notes="proposed new name: TCRBV5S4"
22357..22372
/genes="TCRBV6S6A3N2T"
/notes="conserved 16mer; possible promoter"
join(22475..22523,22646..>22939)
/codon_start=1
/product="TCRBV5S6"
/protein_id="AAB63908.1"

```

Alignment Scores:

Pred. No.:	8.3e+03	Length:	136975
Score:	44.00	Matches:	10
Percent Similarity:	73.33%	Conservative:	1
Best Local Similarity:	66.67%	Mismatches:	4
Query Match:	55.70%	Indels:	0
DB:	9	Gaps:	0

US-09-737-297-3 (1-16) x HSTCRB75A (1-136975)

QY 2 GluglySerThr***AspValTyGlnAsnIleGlnTyzAlaGly 16

Db 122932 GAAGATCTACTCTCAATGTGTATGGCAATATCCAGTCAGCTGGG 122888

RESULT 44

ALI161776
 LOCUS
 DEFINITION Homo sapiens chromosome 6 clone RP11-337K13, 27 unordered pieces.
 ACCESSION ALI161776
 VERSION ALI161776.6 GI:9931705
 KEYWORDS HTG; HTGS PHASE1; HTGS_CANCELLED.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 Sims,S.
 Direct Submission
 AUTHORS
 TITLE
 JOURNAL

Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk

On Aug 28, 2000 this sequence version replaced gi:9863615.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BA337K13
----- Summary Statistics -----
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 137249 bases at least Q40
Consensus quality: 142786 bases at least Q30
Consensus quality: 146178 bases at least Q20
Insert size: 149436; sum-of-contigs
Insert size: 181098; 4.4% error; agarose-ff
Quality coverage: 2.99x in Q20 bases; sum-of-contigs Quality
coverage: 3.07x in Q20 bases; agarose-ff

* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 4296: contig of 4296 bp in length
* 4297 4396: gap of 100 bp
* 4397 12185: contig of 7789 bp in length
* 12186 12285: gap of 100 bp
* 12286 14344: contig of 2059 bp in length
* 14345 14444: gap of 100 bp
* 14445 16466: contig of 2022 bp in length
* 16467 16566: gap of 100 bp
* 16567 21640: contig of 5074 bp in length
* 21641 26015: contig of 4275 bp in length
* 26016 30177: gap of 100 bp
* 30178 30278: contig of 4062 bp in length
* 30279 30593: gap of 100 bp
* 30594 37959: contig of 7682 bp in length
* 37960 38050: gap of 100 bp
* 38051 47429: contig of 9370 bp in length
* 47430 47529: gap of 100 bp
* 47530 52816: contig of 5287 bp in length
* 52817 52916: gap of 100 bp
* 52917 57064: contig of 4148 bp in length
* 57065 57164: gap of 100 bp
* 57165 61171: contig of 4007 bp in length
* 61172 70010: contig of 8739 bp in length
* 70011 70110: gap of 100 bp
* 70111 73274: contig of 3164 bp in length
* 73275 73375: gap of 100 bp
* 73376 80195: contig of 6821 bp in length
* 80196 80295: gap of 100 bp
* 80296 83948: contig of 3653 bp in length
* 83949 84049: gap of 100 bp
* 84050 86450: contig of 2401 bp in length
* 86451 91004: contig of 4455 bp in length
* 91005 91105: gap of 100 bp
* 91106 94864: contig of 3759 bp in length
* 94865 100477: gap of 100 bp
* 100478 100577: contig of 5514 bp in length
* 100578 110898: gap of 100 bp
* 110899 110999: contig of 10321 bp in length
* 111000 114338: gap of 100 bp
* 114339 124405: contig of 9867 bp in length
* 124406 124505: gap of 100 bp
* 124506 129068: contig of 4562 bp in length
* 129069 129167: gap of 100 bp
* 129168 132301: contig of 3134 bp in length
* 132302 132401: gap of 100 bp
* 132402 137230: contig of 4829 bp in length
* 137231 137233: gap of 100 bp
*
* 137331 152036: contig of 14706 bp in length.
* Location/Qualifiers
* 1. 152036
* /organism="Homo sapiens"
* /mol_type="genomic DNA"
* /db_xref="taxon:9606"
* /chromosome="6"
* /clone="RP11-337K13"
* /clone_lib="RPC1-11.2"
* 1. 4296
* /note="assembly fragment:00064"
* /fragment_chain:1"
* 4397 12185
* /note="assembly fragment:01591"
* /fragment_chain:1"
* 12286 14344
* /note="assembly fragment:01596"
* /fragment_chain:1"
* 14445 16466
* /note="assembly fragment:00152"
* /fragment_chain:2"
* 16567 21640
* /note="assembly fragment:01276"
* /fragment_chain:2"
* 21741 26015
* /note="assembly fragment:00919"
* /fragment_chain:2"
* 26116 30177
* /note="assembly fragment:00311"
* /fragment_chain:3"
* 30278 37959
* /note="assembly fragment:00504"
* /fragment_chain:3"
* 38060 47429
* /note="assembly fragment:01383"
* /fragment_chain:3"
* 47530 52816
* /note="assembly fragment:01092"
* /fragment_chain:4"
* 52917 57064
* /note="assembly fragment:01674"
* /fragment_chain:4"
* 57165 61171
* /note="assembly fragment:01031"
* /fragment_chain:4"
* 61272 70010
* /note="assembly fragment:00275"
* /fragment_chain:5"
* 70111 73274
* /note="assembly fragment:01563"
* /fragment_chain:5"
* 73375 80195
* /note="assembly fragment:00326"
* /fragment_chain:6"
* 80296 83948
* /note="assembly fragment:01722"
* /fragment_chain:6"
* 84049 86449
* /note="assembly fragment:00994"
* /fragment_chain:7"
* 86550 91004
* /note="assembly fragment:01306"
* /fragment_chain:7"
* 91105 94863
* /note="assembly fragment:01540"
* /fragment_chain:8"
* 94964 100477
* /note="assembly fragment:00335"
* /fragment_chain:8"
* 100578 110898
* /note="assembly fragment:00528"
* /fragment_chain:8"
* 110999 114438
* /note="assembly fragment:01077"

```

misc_feature 114539..124405
/note="assembly_fragment:01106"
misc_feature 124506..129067
/note="assembly_fragment:01514"
misc_feature 129168..132301
/note="assembly_fragment:01598"
misc_feature 132402..137230
/note="assembly_fragment:01664"
misc_feature 137331..152036
/note="assembly_fragment:00343
clone_end:T7
vector_side:right"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 9.33e+03 Length: 152036
Score: 44.00 Matches: 7
Percent Similarity: 83.33% Conservative: 3
Best Local Similarity: 58.33% Mismatches: 2
Query Match: 55.70% Indels: 0
DB: 2 Gaps: 0

```

US-09-737-297-3 (1-16) x AL161776 (1-152036)

Cy 5 Thr***AspValTyrGlnAsnIleGIntYrAlaGly 16

Db 97897 ACITGGACATTTATCAAAATTTATCATATGCTGCT 97932

RESULT 45

LOCUS AC134321/c 159314 bp DNA linear HTG 23-NOV-2002

DEFINITION Felis catus clone RP86-48B21, WORKING DRAFT SEQUENCE, 6 ordered

pieces.

AC134321

VERSION AC134321.2 GI:25188949

KEYWORDS HTG; HTGS PHASE2; HTGS_DRAFT.

SOURCE Felis catus (cat)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.

1 (bases 1 to 159314)

Author, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,

Benjamin, B., Blakeley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,

Carriaga, K., Coleman, B., Engle, J., Granite, S., Guan, X., Gupta, J.,

Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E.,

Laric, P., Lee-Jin, S.-Q., Legaspi, R., Maduro, Q.L., McDowell, J.,

Margulies, E.H., Masiello, C., Maskeri, B., McDowell, J.,

Paguirigan, C., Pearson, R., Portnoy, M.E., Prasad, A.,

Reddix-Dugue, N., Schandler, K., Schueler, M.G., Sison, C.,

Stantripop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W., Vogt, J.L.,

Wetherby, K.D., Wiggins, L., Young, A., and Green, E.D.

NTSC Comparative Sequencing Initiative

Unpublished

2 (bases 1 to 159314)

Green, E.D.

Direct Submission

Submitted (25-SEP-2002) NIH Intramural Sequencing Center, 8717

GroveMont Circle, Gaithersburg, MD 20877, USA

3 (bases 1 to 159314)

Green, E.D.

Direct Submission

Submitted (23-NOV-2002) NIH Intramural Sequencing Center, 8717

GroveMont Circle, Gaithersburg, MD 20877, USA

On Nov 23, 2002 this sequence version replaced gi:23308032.

----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: nisc_zoo@hgri.nih.gov

----- Project Information

Center project name: csw

Center clone name: 048521

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 158227 bases at least Q40
 Consensus quality: 158628 bases at least Q30
 Consensus quality: 158772 bases at least Q20
 Insert size: 149000; agarose-fp
 Insert size: 158814; sum-of-contigs
 Quality coverage: 12.07x in Q20 bases; agarose-fp
 Quality coverage: 11.32x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

* 1 33762: contig of 33762 bp in length
 * 33763 33862: gap of unknown length
 * 33863 42899: contig of 9037 bp in length
 * 42900 42999: gap of unknown length
 * 43000 45209: contig of 2210 bp in length
 * 45310 45309: gap of unknown length
 * 45310 48354: contig of 3045 bp in length
 * 48355 48454: gap of unknown length
 * 48455 50966: contig of 2512 bp in length
 * 50967 51066: gap of unknown length
 * 51067 159314: contig of 108248 bp in length.

FEATURES

source

1..159314
 /organism="Felis catus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9685"
 /clone_lib="RP86-48B21"
 /clone_lib="RP86"

misc_feature

1..33762
 /note="assembly_fragment
 clone_end:SP6
 vector_side:left"

misc_feature

33863..42899
 /note="assembly_fragment"
 /db_xref="taxon:9685"

misc_feature

43000..45209
 /note="assembly_fragment"

misc_feature

45310..48354
 /note="assembly_fragment"

misc_feature

48455..50966
 /note="assembly_fragment"

misc_feature

51067..159314
 /note="assembly_fragment
 clone_end:T7
 vector_side:right"

ORIGIN

Alignment Scores:

Pred. No.: 9.83e+03 Length: 159314

Score: 44.00 Matches: 8

Percent Similarity: 66.67% Conservative: 2

Best Local Similarity: 53.33%
 Query Match: 55.70%
 DB: 2

DB: 10 Gaps: 0
 US-09-737-297-3 (1-16) x AL591712 (1-168958)

US-09-737-297-3 (1-16) x AC134321 (1-159314)

Qy 1 AlaGluGlySerThr***AspValTyrGlnAsnleGlnTyrAlaGly 15
 |||:|||||
 Db 57784 GCCAAGGTAAACACAGACATTTATACCATAGTTGGTATGCC 57740

RESULT 46
 AL591712 168958 bp DNA linear ROD 07-NOV-2002
 LOCUS Mouse DNA sequence from clone RP23-138C10 on chromosome 2, complete
 DEFINITION
 ACCESSION AL591712
 VERSION AL591712
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 168958)
 REFERENCE
 AUTHORS Kay, M.
 TITLE Direct Submission
 JOURNAL Submitted (10-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Oct 13, 2002 this sequence version replaced gi:23894359.
 COMMENT
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-138C10 is from the RPCT-23 Mouse PAC library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: PBACe3.6.

FEATURES
 source
 1. 168958
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="2"
 /clone="RP23-138C10"
 /clone_lib="RPCI-23"

ORIGIN

Alignment Scores:
 Pred. No.: 1.05e+04 Length: 168958
 Score: 44.00 Matches: 8
 Percent Similarity: 52.50% Conservative: 2
 Best Local Similarity: 50.00% Mismatches: 6
 Query Match: 55.70% Indels: 0

LOCUS AL356421 170532 bp DNA linear PRI 30-SEP-2000
 DEFINITION Human DNA sequence from clone RP11-550C4 on chromosome 6, complete
 sequence.
 ACCESSION AL356421
 VERSION AL356421.10 GI:10443437
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 170532)
 REFERENCE
 AUTHORS Corby, N.
 TITLE Direct Submission
 JOURNAL Submitted (29-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk

On Oct 1, 2000 this sequence version replaced gi:10186530.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6> RP11-550C4 is from the library RPCI-11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR: pbACe3.6>

IMPORTANT: This sequence is not the entire insert of clone RP11-550C4. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP11-550C4 is at 1 in this sequence. The true left end of clone RP3-402H5 is at 170433 in this sequence. The true right end of clone RP11-81210 is at 111382 in this sequence.

FEATURES
 source
 1. 170532
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone="RP11-550C4"
 /clone_lib="RPCI-11.2"

misc_feature

41769
 /note="Tandem repeat. Forced join. Gap size estimated to be approximately 150bp by EcoRI and HindIII restriction enzyme digest data."

ORIGIN

Alignment Scores:

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 215422)
 AUTHORS Slightom, J.L., Siemieniuk, D.R., Siew, L.C., Koop, B.F. and Hood, L.
 TITLE Nucleotide sequence analysis of 77.7 kb of the human V beta T-cell
 receptor gene locus: direct primer-walking using cosmid template
 DNAs
 JOURNAL Genomics 20 (2), 149-168 (1994)
 MEDLINE 94292194
 PUBMED 8020962
 REFERENCE 2 (bases 1 to 215422)
 AUTHORS Rowen, L., Koop, B.F. and Hood, L.
 TITLE The complete 685-kilobase DNA sequence of the human beta T cell
 receptor locus
 JOURNAL Science 272 (5269), 1755-1762 (1996)
 MEDLINE 96256474
 PUBMED 8650574
 REFERENCE 3 (bases 1 to 215422)
 AUTHORS Rowen, L., Seto, J., Smit, A., Acharya, C., Ahearn, M.E., Ankener, M.,
 Baskin, D., Bumgarner, R., Chen, L., Chen, N., Deshpande, P., Faust, J.,
 Howard, S., Jerome, N., Koop, B.F., Lee, H., Loretz, C., Paepker, B.,
 Zackrone, K. and Hood, L.
 TITLE Sequence determination of the human T cell receptor beta locus:
 Strategy and error analysis
 JOURNAL Unpublished
 REFERENCE 4 (bases 1 to 215422)
 AUTHORS Rowen, L., Wang, K., Boyesen, C., Ahearn, M.E., Charnley, P., Paepker, B.,
 Lee, I., Chen, L., Trask, B., Nickerson, D., Seto, D. and Hood, L.
 TITLE Sequence variation among several haplotypes in the human T cell
 receptor beta locus
 JOURNAL Unpublished
 REFERENCE 5 (bases 1 to 215422)
 AUTHORS Rowen, L.
 TITLE Direct Submission
 JOURNAL Submitted (15-OCT-1994) L.Rowen leerowen@u.washington.edu
 REFERENCE 6 (bases 1 to 215422)
 AUTHORS Rowen, L.
 TITLE Direct Submission
 JOURNAL Submitted (25-JUN-1997)
 COMMENT This sequence overlaps section 1 of the human T cell receptor beta
 locus, GenBank Accession Number U66059, by 9647 bases. It overlaps
 section 3, GenBank Accession Number U66061, by 20616 bases.

FEATURES
 source
 1..215422
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /map="7q35"
 <1..9638
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="G15"
 /cell_line="ATCC1475"
 /clone_lib="Kai Wang's"
 /germline
 /note="(vector PWE15A)"
 source
 1..39055
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="X1A"
 /cell_line="ATCC1475"
 /clone_lib="Kai Wang's"
 /germline
 /note="(vector PWE15A)"
 source
 32875..65622
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="A27"

/cell_line="CGM1, haplotype A"
 /clone_lib="YAC 234 A6F6"
 /germline
 /note="(vector PWE15A)"
 63647..76133
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="A212partial"
 /cell_line="CGM1, haplotype A"
 /clone_lib="YAC 234 A6F6"
 /germline
 /note="(vector PWE15A)"
 75206..110547
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="A14"
 /cell_line="CGM1, haplotype A"
 /clone_lib="YAC 234 A6F6"
 /germline
 /note="(vector PWE15A)"
 106834..145015
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="H7.1"
 /germline
 123850..163388
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="H12.18"
 /germline
 146622..184576
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="H130"
 /germline
 181186..215422
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="A18"
 /cell_line="CGM1, haplotype A"
 /clone_lib="YAC 234 A6F6"
 /germline
 /note="(vector PWE15A)"
 194806..215422
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="C215"
 /cell_line="CGM1 haplotype A"
 /clone_lib="YAC 234 A72B3"
 /germline
 /note="(vector PWE15A)"
 1..134786
 /note="Overlap with a sequence contig derived from cosmids
 7-5A, 4-8, 2-9, CA33, and K34, isolated from cell line
 CGM1, haplotype B."
 563
 /note="cosmids G15 and X1A: a; CGM1 haplotype B: t"
 /replace="t"
 1660..1751
 /note="(TG)46, 86% conserved; putative"
 /rpt_type=tandem
 1691
 /note="cosmids G15 and X1A: g; CGM1 haplotype B: gtagag"
 /replace="gtgag"
 2050..38396
 /note="Overlap with cosmid B27, isolated from CGM1 cell


```

line, haplotype A."
2186
/notes="cosmids G15 and X1A: t; CGM1 haplotypes B and A: c"
/replace="c"
2715
/notes="cosmids G15 and X1A: c; CGM1 haplotypes B and A: t"
/replace="t"
2796
/notes="cosmids G15 and X1A: a; CGM1 haplotypes B and A: g"
/replace="g"
3071..3309
/notes="LINE; putative"
/tp family="LIMA4"
3340
/notes="cosmids G15 and X1A: c; CGM1 haplotypes B and A: a"
/replace="a"
3429
/notes="LINE; putative"
/tp family="LIMB1; fragment 1"
4294..4597
/notes="putative"
/tp family="Alu"
4353
/notes="cosmids G15 and X1A: g; CGM1 haplotypes B and A: a"
/replace="a"
4481
/notes="cosmids G15 and X1A: a; CGM1 haplotypes B and A: g"
/replace="g"
4493..4494
/notes="cosmids G15 and X1A: aa; CGM1 haplotypes B and A: cc"
/replace="cc"
4571..4575
/notes="cosmids G15 and X1A: aaaa; CGM1 haplotypes B and A: a"
/replace="a"
4731
/notes="LINE; putative"

```

```

Alignment Scores:
Pred. No.: 1.38e+04 Length: 215422
Score: 44.00 Matches: 10
Percent Similarity: 73.33% Conservative: 1
Best Local Similarity: 66.67% Mismatches: 4
Query Match: 55.70% Indels: 0
DB: 9 Gaps: 0

```

US-09-737-297-3 (1-16) x U66060 (1-215422)

QY 2 GluglySeThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16

Db 120742 GAAGATCTACTCTCAATGTATGCAATATCCAGTCAGCTGG 120698

```

RESULT 50
AC112452/c 224429 bp DNA linear HTG 20-NOV-2002
LOCUS Rattus norvegicus clone CH230-56P13, WORKING DRAFT SEQUENCE, 3
DEFINITION
unordered pieces.
AC112452
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 224429)

```

```

REFERENCE
AUTHORS Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

```

```

Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleaveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Georgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpaty, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, Y., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuwa, L., Loulseg, H., Lozada, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwackelme, O., Okwuonu, G., Olarunpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Register, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rivas, C., Rokey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, M., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shivatsbeyn, A., Sisson, I., Sitter, C.D., Smales, D.,
Sned, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Staimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 224429)
Worley, K.C.
Direct Submission
Submitted (21-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 224429)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:23265975.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM

```

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center Project name: GRZG

Center clone name: CH230-56P13

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 218820 bases at least Q40

Consensus quality: 220372 bases at least Q30

Consensus quality: 221633 bases at least Q20

Estimated insert size: 225534; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 3 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 124621: contig of 124621 bp in length

* 124622 124721: gap of unknown length

* 124722 223259: contig of 98538 bp in length

* 223260 224429: gap of unknown length

* 224429 224429: contig of 1070 bp in length.

Location/Qualifiers

1. .224429

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-56P13"

complement(94150..94821)

/note="clone boundary"

clone_end:Sp6

site:

end_sequence: BH295752"

ORIGIN

Alignment Scores:	1.45e+04	Length:	224429
Pred. No.:	44.00	Matches:	8
Score:	75.00%	Conservative:	4
Percent Similarity:	50.00%	Mismatches:	4
Best Local Similarity:	55.70%	Indels:	0
Query Match:	2	Gaps:	0
DB:			

US-09-737-297-3 (1-16) x AC112452 (1-224429)

QY 1 AlaGluGlySerThr***AspValtyrGlnAsnIleGlnTyrrAlaGly 16

DB 7757 AGTGAGGGTAATAAAGCGCGACCTATTCCAGAACAGCAGCGCGGT 7710

Search completed: March 1, 2004, 11:00:50

Job time : 2267 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - nucleic search, using frame_plus_p2n model

run on: March 1, 2004, 10:14:14 ; Search time 342 seconds
(without alignments)
198.746 Million cell updates/sec

title: US-09-737-297-3

perfect score: 79

sequence: 1 AEGSYXDVYQNTQYAG 16

scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US09737297/runat_01032004_085322_3450/app_query.fasta_1.199
-DB=N Geneseq 29Jan04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=150 -DOCALLIGN=200 -THR SCORE=100 -THR MIN=0 -ALIGN=50
-MODE=LOCAL -OUTFMT=pro -NORM=ext -HEASIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09737297@cgn_1_708 @runat_01032004_085322_3450 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq 29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1980s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004as:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	59.5	4216	5	Aas76703 DNA encod
2	46	58.2	580	3	Aaf09054 Fusarium
3	43	54.4	110000	6	Continuation (7 of
4	42	53.2	724	7	ACA9172_06
5	42	53.2	1140	7	ABX06830 S. pneumo
6	42	53.2	1195	2	Aaz28864 Streptomy
7	42	53.2	1195	2	Aaz28865 Streptomy
8	42	53.2	5329	3	AAa88900 Mouse lam

9	42	53.2	5329	3	AAC83712	Mouse lam
10	42	53.2	5329	6	ABQ72911	Mouse lam
11	42	53.2	5689	3	AAA88899	Mouse lam
12	42	53.2	5689	6	ABQ72911	Mouse lam
13	42	53.2	5689	6	ABQ72910	Mouse lam
14	42	53.2	11384	2	AAV52178	Streptococ
15	42	53.2	110000	7	ABS56454_09	Continuation (10 o
16	41	51.9	125	6	ABN62097	Human can
17	41	51.9	441	5	ABV46101	Human pro
18	41	51.9	445	6	ABN76334	Human ORF
19	41	51.9	675	2	AAQ36563	Endo-xyla
20	41	51.9	799	7	AAQ36563	Woody pla
21	41	51.9	1078	5	AA577451	DNA encod
22	41	51.9	1230	4	AA553536	Haemophil
23	41	51.9	1626	5	AA573157	DNA encod
24	41	51.9	2010	9	ADC68525	Lolium pe
25	41	51.9	2139	6	ABK98515	CDNA sequ
26	41	51.9	4683	5	AA569826	DNA encod
27	41	51.9	8301	4	ABL23219	Drosophil
28	41	51.9	11142	4	ABL23218	Drosophil
29	41	51.9	43360	8	ACD19193	E. coli
30	41	51.9	43450	9	ADC00771	Enterobac
31	41	51.9	45325	8	ACD19247	E. coli
32	41	51.9	110000	2	AA742063_17	Continuation (18 o
33	40	50.6	415	7	ABX46781	Bovine ES
34	40	50.6	517	4	AAH08199	Human CDN
35	40	50.6	1104	3	AAAD01958	Murine TH
36	40	50.6	1371	4	AA553136	Enterococ
37	40	50.6	1554	7	ACD66743	Secrete
38	40	50.6	1773	6	AA597277	Neisseria
39	40	50.6	1773	6	AA597234	Neisseria
40	40	50.6	1773	6	AAI67435	N. mening
41	40	50.6	1776	7	ACA41645	Prokaryot
42	40	50.6	1793	3	AAFL6209	Human pro
43	40	50.6	1812	4	AAH22890	R. ruber
44	40	50.6	1927	4	AAH18171	Human CDN
45	40	50.6	2019	3	AA606675	Moloney m
46	40	50.6	2151	6	ABK13201	DNA encod
47	40	50.6	2216	3	AAZ33321	Human sec
48	40	50.6	2467	5	AA501368	Murine TA
49	40	50.6	2467	7	ACD66742	Secrete
50	40	50.6	3303	4	ABL29911	Drosophil
51	40	50.6	3691	7	ABX63796	Human CDN
52	40	50.6	4871	6	ABZ11566	Human pol
53	40	50.6	4983	8	ABZ34854	Coding se
54	40	50.6	5849	8	ADA03030	Mouse MCG
55	40	50.6	5849	9	ADB72768	Mouse MCG
56	40	50.6	5849	9	ADC85510	Mouse MCG
57	40	50.6	20978	4	ABL20786	Drosophil
58	40	50.6	49767	3	AAAB1458	N. mening
59	40	50.6	90091	9	ADC85509	Mouse MCG
60	40	50.6	90100	8	ADA03029	Mouse MCG
61	40	50.6	90100	9	ADB72767	Mouse MCG
62	40	50.6	110000	3	AAA81489_7	Continuation (8 of
63	40	50.6	172325	3	AA216133	Neisseria
64	39.5	50.0	180	7	ABX07165	S. pneumo
65	39.5	50.0	1766	2	AAV52368	Streptococ
66	39.5	50.0	110000	7	ABS56454_12	Continuation (13 o
67	39	49.4	45	2	AAQ04545	T90 Guess
68	39	49.4	45	2	AAI59384	Residual
69	39	49.4	50	6	ABZ06254	Human pro
70	39	49.4	577	5	ABV59285	Human leu
71	39	49.4	939	4	AAQ03005	Bacillus
72	39	49.4	998	2	AAQ04548	Residual
73	39	49.4	998	2	AAI63356	Bacillus
74	39	49.4	998	2	AAI59387	RP-I gene
75	39	49.4	1137	6	ABN70926	Streptococ
76	39	49.4	1155	6	ABN68225	Streptococ
77	39	49.4	1299	5	AA566458	DNA encod
78	39	49.4	1302	3	AA59422	Human sec
79	39	49.4	1302	7	ABT16805	Human sec
80	39	49.4	1302	7	ABZ67033	Human sec
81	39	49.4	1302	9	ADC20154	Human sec

C 82	39	49.4	1492	4	ABL11047
C 83	39	49.4	1501	6	ABV99610 Methionin
C 84	39	49.4	1772	6	ABL52763 Plant def
C 85	39	49.4	1872	6	ABQ68034 Listeria
C 86	39	49.4	1872	6	ABQ69949 Listeria
C 87	39	49.4	1872	6	ACA36328 Prokaryot
C 88	39	49.4	2000	6	ABZ15581 Arabidops
C 89	39	49.4	2000	6	ABZ15243 Arabidops
C 90	39	49.4	2466	2	AAQ58336 S. cerevi
C 91	39	49.4	2491	4	AAH17424 Human CDN
C 92	39	49.4	2598	7	ACA26426 Prokaryot
C 93	39	49.4	2623	4	ABL11407 Drosophil
C 94	39	49.4	2905	6	AAK99305 CDNA enco
C 95	39	49.4	3072	4	AAK69732 Human lmm
C 96	39	49.4	3089	2	AAQ58337 S. cerevi
C 97	39	49.4	3378	3	AAF21917 Human bre
C 98	39	49.4	3783	8	ADA31333 DNA enco
C 99	39	49.4	3805	4	ABL11046 Drosophil
C 100	39	49.4	3907	6	ABQ54325 Human ova
C 101	39	49.4	4813	7	ABT16935 Human sec
C 102	39	49.4	4813	7	ABZ67594 Human sec
C 103	39	49.4	4813	9	ADC20738 Human sec
C 104	39	49.4	4921	4	ABL29358 Drosophil
C 105	39	49.4	5456	4	ABL11406 Drosophil
C 106	39	49.4	8278	4	AAK99138 Canine re
C 107	39	49.4	11391	3	AAA90046 Bovine ly
C 108	39	49.4	11391	3	AAA88306 Bovine ly
C 109	39	49.4	23378	4	AAK9555 Human dig
C 110	39	49.4	23378	4	AAK9555 Human dig
C 111	39	49.4	23378	8	ADB60783 Connectiv
C 112	39	49.4	25988	4	ABL09406 Drosophil
C 113	39	49.4	42048	4	AAK71918 Human lmm
C 114	39	49.4	110000	6	ABN71527 08
C 115	39	49.4	110000	6	ABX08336 10
C 116	39	49.4	110000	6	ABA03041 11
C 117	39	49.4	185371	6	ABT10718 Human bre
C 118	39	49.4	326014	6	ABK89296 Human gen
C 119	38.5	48.7	409	3	AAK22236 Human sec
C 120	38.5	48.7	577	9	ADD52220 Mouse mlt
C 121	38.5	48.7	732	4	AAH06769 Human CDN
C 122	38.5	48.7	979	4	AAH161269 Human pol
C 123	38.5	48.7	1449	5	AAK94543 DNA enco
C 124	38.5	48.7	2697	4	AAH14502 Human CDN
C 125	38.5	48.7	2707	5	ABX71417 Intracell
C 126	38.5	48.7	2707	7	ACF34471 Gene enco
C 127	38.5	48.7	2748	4	AAK26861 Human CDN
C 128	38.5	48.7	2821	4	AAK59483 Human pol
C 129	38.5	48.7	2940	3	AAK77535 Human ORF
C 130	38.5	48.7	3246	2	AAK99525 Polynucle
C 131	38.5	48.7	20272	4	AAK03527 Human rep
C 132	38.5	48.7	20272	5	ABK72118 Human ova
C 133	38.5	48.7	20272	6	ABK91710 Human ova
C 134	38.5	48.7	25619	4	AAK26966 Human gen
C 135	38.5	48.7	25619	4	AAK03525 Human rep
C 136	38.5	48.7	25619	5	ABK72116 Human ova
C 137	38.5	48.7	25619	6	ABK91708 Human ova
C 138	38	48.1	196	2	AAK12271 Human bia
C 139	38	48.1	283	3	AAK46655 Zea may
C 140	38	48.1	318	4	AAI20416 Probe #10
C 141	38	48.1	318	4	ABA65454 Human toe
C 142	38	48.1	318	4	AAI45620 Probe #14
C 143	38	48.1	318	4	ABA47561 Human bre
C 144	38	48.1	318	4	ABA32551 Probe #11
C 145	38	48.1	318	4	AAK39606 Human bon
C 146	38	48.1	318	4	AAK13862 Human bra
C 147	38	48.1	318	4	ABS39196 Human liv
C 148	38	48.1	318	5	AAI06116 Probe #61
C 149	38	48.1	318	6	ABS13697 Human gen
C 150	38	48.1	322	5	AAK94541 DNA enco

ALIGNMENTS

RESULT 1	
AAK76703	
ID	AAK76703 standard; cDNA; 4216 BP.
XX	AAK76703
AC	AAK76703
XX	13-FEB-2002 (first entry)
DT	
XX	DNA encoding novel human diagnostic protein #12507.
DE	Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	Homo sapiens.
OS	
XX	WO200175067-A2.
PN	
XX	11-OCT-2001.
XX	
XX	30-MAR-2001; 2001WO-US008631.
XX	
XX	31-MAR-2000; 2000US-00540217.
XX	23-AUG-2000; 2000US-00649167.
XX	(HYPSE-) HYSEQ INC.
XX	Dmanac RT, Liu C, Tang YT;
XX	WPI: 2001-639362/73.
XX	P-PSDB; ABG12516.
XX	
XX	New isolated polynucleotide and encoded polypeptides, useful in
XX	PT diagnostics, forensics, gene mapping, identification of mutations
XX	PT responsible for genetic disorders or other traits and to assess
XX	PT biodiversity.
XX	
XX	Claim 1; SEQ ID NO 12507; 103pp; English.
XX	
XX	The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX	sequences. (I) is useful as hybridisation probes, polymerase chain
XX	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX	and in recombinant production of (II). The polynucleotides are also used
XX	in diagnostics as expressed sequence tags for identifying expressed
XX	genes. (II) is useful in gene therapy techniques to restore normal
XX	activity of (II) or to treat disease states involving (II). (II) is
XX	useful for generating antibodies against it, detecting or quantitating a
XX	polypeptide in tissue, as molecular weight markers and as a food
XX	supplement. (II) and its binding partners are useful in medical imaging
XX	of sites expressing (II). (I) and (II) are useful for treating disorders
XX	CC involving aberrant protein expression or biological activity. The
XX	CC polypeptide and polynucleotide sequences have applications in
XX	CC diagnostics, forensics, gene mapping, identification of mutations
XX	CC and to produce other types of data and products dependent on DNA and
XX	CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
XX	CC coding sequences of the invention. Note: The sequence data for this
XX	CC patent did not appear in the printed specification, but was obtained in
XX	CC ftp.wipo.int/pub/published_pct_sequences
XX	XX Sequence 4216 BP; 1555 A; 792 C; 830 G; 1038 T; 0 U; 1 Other;
XX	
XX	Alignment Scores:
Pred. No.:	23 Length: 4216
Score:	47.00 Matches: 8
Percent Similarity:	73.3% Conservatives: 3
Best Local Similarity:	53.3% Mismatches: 4
Query Match:	59.4% Indels: 0
DB:	5 Gaps: 0
US-09-737-297-3 (1-16) x AAK76703 (1-4216)	
QY	2 GluGlySerThr***AspValtyrGlnAsnIleGlnTyrAlaGly 16

PF 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

DR WPI; 2003-029926/02.

DR P-PSDB; ABU45302.

XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 37042; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target

CC prokaryotic essential genes. Note: The sequence data for this patent did

CC not form part of the printed specification, but was obtained in

CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 724 BP; 181 A; 160 C; 192 G; 191 T; 0 U; 0 Other;

XX Alignment Scores:

XX Pred. No.: 30.7 Length: 724

XX Score: 42.00 Matches: 7

XX Percent Similarity: 69.23% Conservative: 2

XX Best Local Similarity: 53.85% Mismatches: 4

XX Query Match: 53.16% Indels: 0

XX DB: 7 Gaps: 0

XX US-09-737-297-3 (1-16) x ACH49172 (1-724)

QY 2 GlucylserThr***AspValTyrGlnAsnIleGlnTyr 14

DB 57 GATCGCGCACAGAGATGTTTACTTACTGATTCGAATAT 95

RESULT 5

ID ABX06830 standard; DNA; 1140 BP.

XX AC ABX06830;

XX 27-OCT-2003 (revised)

DT 11-FEB-2003 (first entry)

XX S. pneumoniae type 4 strain coding region #1118.

XX Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;

XX ear infection; antiinflammatory; antibacterial; immunostimulant;

XX auditory; respiratory; gene therapy; vaccine.

XX Streptococcus pneumoniae; type 4 strain.

OS WO200277021-A2.

PN 03-OCT-2002.

PD 27-MAR-2002; 2002WO-IB002163.

PP 27-MAR-2001; 2001GB-00007658.

XX (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

XX Masignani V, Tettelin H, Fraser C;

PI WPI; 2003-040579/03.

XX P-PSDB; ABU01542.

XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,

XX useful as medicaments for treating or preventing a disease or infection

XX due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or

XX ear infection.

XX Claim 6; SEQ ID NO 2235; 56pp; English.

XX The invention relates to a protein comprising or having at least 50%

XX identity to any of the 2469 amino acid sequences, identified in the

XX specification (available on a computer readable format), or its fragment,

XX expressed from 2469 of 2489 identified DNA coding regions from the

XX Streptococcus pneumoniae type 4 strain genomic sequence appearing as

XX ABS56454. Also included are an antibody which binds one of the proteins,

XX treating a patient by administering the protein, DNA or antibody (in a

XX composition), a kit comprising first and second primers, which are the

XX nucleic acid cited above or fragments between nucleotides 8-100 of a

XX sequence not defined in the specification, for amplifying a target

XX sequence contained within a Streptococcus nucleic acid sequence, where

XX the first primer is substantially complementary to the target sequence

XX and the second primer is substantially complementary to the complement of

XX the target sequence, and where the parts of the primers having

XX substantial complementarity define the termini of the target sequence to

XX be amplified, assay comprising contacting a test compound with the

XX protein, and determining whether the test compound binds to the protein

XX and a Streptococcus pneumoniae bacterium, where one or more genes

XX encoding the proteins has been rendered inactive. The proteins, nucleic

XX acid molecules, antibody and compositions are useful as medicaments for

XX treating or preventing a disease or infection due to streptococcus

XX bacteria, particularly *S. pneumoniae*, such as pneumonia, sepsis, otitis

XX media or ear infection. They are also useful in developing vaccines,

XX diagnostics and antibiotics. The methods are useful for identifying

XX immunodominant proteins. The present sequence is one of the 2489

XX identified coding regions from the genomic sequence. Note: The sequence

XX data for this patent did not form part of the printed specification, but

XX was obtained in electronic format directly from WIPO at

XX ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to

XX standardise OS field)

XX SQ Sequence 1140 BP; 337 A; 204 C; 265 G; 334 T; 0 U; 0 Other;

XX Alignment Scores:

XX Pred. No.: 53.7 Length: 1140

XX Score: 42.00 Matches: 7

db 742 AGCAGCGGCGACCTACGAGCTTACACAGACGCGGTAC 783

RESULT 8

ID AAA88900 standard; cDNA; 5329 BP.

AC AAA88900;

XX 05-MAR-2001 (first entry)

XX Mouse laminin 2 beta-1 chain (mature region) cDNA.

XX Laminin 2; mouse; nerve regeneration; angiogenic; cell adhesion;

XX degenerative muscle disorder; muscular dystrophy; cell therapy; ss.

XX Mus musculus.

XX Key Location/Qualifiers

XX mat_peptide 1..5175

XX /*tag= a

XX WO200066730-A2.

XX 09-NOV-2000.

XX 28-APR-2000; 2000WO-US011378.

XX 30-APR-1999; 99US-0131720P.

XX 15-JUN-1999; 99US-0139198P.

XX 12-JUL-1999; 99US-0143289P.

XX 24-SEP-1999; 99US-0155945P.

XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX Yurchenco P;

XX WPI; 2000-687537/67.

XX P-PSDB; AAB19800.

XX Purified laminin 2 protein, useful for research and therapeutic purposes

XX including peripheral nerve regeneration, treatment of degenerative muscle

XX disorders, angiogenesis regulation, and ex vivo cell therapy.

XX Claim 4; Page 219-226; 305pp; English.

XX The present sequence is that of cDNA encoding the mature beta-1 chain

XX (see AAB19800) of mouse laminin 2. Laminin 2 is composed of alpha-2 (400

XX kDa), beta-1 (100 kDa) and gamma-1 (100 kDa) chains. It is thought to be

XX specifically required for stabilizing myotubes during skeletal muscle

XX development, and for preventing apoptosis. Genetic defects in human

XX laminin 2 structure or expression are associated with a major type of

XX congenital muscular dystrophy. Laminin 2 is also thought to be important

XX in Schwann cell/basal lamina interactions. The invention provides laminin

XX 2 alpha-2, beta-1 and gamma-1 chain polypeptides (see AAB19791-806) and

XX the polynucleotides encoding them (see AAA8891-906), methods for making

XX recombinant laminin 2, cells that express recombinant laminin 2, and

XX methods for using purified laminin 2 for research and therapeutic

XX purposes including peripheral nerve regeneration, treatment of

XX degenerative muscle disorders, angiogenesis regulation, promoting cell

XX attachment and migration, ex vivo cell therapy, improving the take of

XX grafts, improving the biocompatibility of medical devices and preparing

XX improved culture devices and media

XX Sequence 5329 BP; 1397 A; 1284 C; 1472 G; 1176 T; 0 U; 0 Other;

XX Alignment Scores:

XX Pred. No.: 358 Length: 5329

XX Score: 42.00 Matches: 7

XX Percent Similarity: 71.43% Conservative: 3

XX Best Local Similarity: 50.00% Mismatches: 4

XX Query Match: 53.16% Indels: 0

XX DB: 3 Gaps: 0

US-09-737-297-3 (1-16) x AAA88900 (1-5329)

3 GlySerThr***AspValTyrGlnAsnIleClnTyrAlaGly 16

3723 GCACAGACGCTGGAGTTTATCAAAACTCGATATCAGG 3764

RESULT 9

AAC83712

ID AAC83712 standard; cDNA; 5329 BP.

XX AAC83712;

XX 02-MAR-2001 (first entry)

XX Mouse laminin 8 cDNA, SEQ ID NO: 19.

XX Mouse; laminin 8; neuroprotective; angiogenic; osteopathic;

XX antiarteriosclerotic; glycoprotein; mesenchymal tissue injury;

XX vascular tissue injury; neural injury; angiogenesis regulation; ss.

XX Mus musculus.

XX WO200066732-A2.

XX 09-NOV-2000.

XX 28-APR-2000; 2000WO-US011543.

XX 30-APR-1999; 99US-0131720P.

XX 21-AUG-1999; 99US-0149738P.

XX 24-SEP-1999; 99US-0155945P.

XX 11-FEB-2000; 2000US-0182012P.

XX (BIOS-) BIOSTRATUM INC.

XX Korntesmaa J, Tryggvason K;

XX WPI; 2000-687539/67.

XX P-PSDB; AAB48451.

XX Purified laminin 8 protein, useful for research and therapeutic purposes

XX including peripheral nerve regeneration, treatment of degenerative muscle

XX disorders, angiogenesis regulation, and ex vivo cell therapy.

XX Claim 4; Page 182-189; 245pp; English.

XX The present sequence encodes a laminin 8 polypeptide chain. Laminins are

XX a family of heterotrimeric glycoproteins that function via binding

XX interactions with neighboring cell receptors and by forming laminin

XX networks. They are signaling molecules which influence cellular

XX function. Laminin 8 is useful for treating injuries to tissue of

XX mesenchymal origin, such as bone, cartilage, tendon, and ligament,

XX treating injuries to vascular tissue, promoting cell attachment and

XX migration, ex vivo cell therapy, improving the biocompatibility of

XX medical devices, and preparing improved cell culture devices and media.

XX Laminin 8 is also useful for promoting re-endothelialisation at the site

XX of vascular injuries, improving the take of grafts, improving the

XX biocompatibility of medical devices, treating neural injuries (neural

XX regeneration), regulating angiogenesis, and promoting cell attachment and

XX migration

XX Sequence 5329 BP; 1397 A; 1284 C; 1472 G; 1176 T; 0 U; 0 Other;

XX Alignment Scores:

XX Pred. No.: 358 Length: 5329

XX Score: 42.00 Matches: 7

XX Percent Similarity: 71.43% Conservative: 3

XX Best Local Similarity: 50.00% Mismatches: 4

XX Query Match: 53.16% Indels: 0

XX DB: 3 Gaps: 0

US-09-737-297-3 (1-16) x AAC83712 (1-5329)

3 GlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16
|||||
3723 GGCAGACAGCTGGAGTTTATCAAAACTCCGATATTCAGGG 3764

RESULT 10

BQ72911

D ABO72911 standard; cDNA; 5329 BP.

X C ABO72911;

X T 19-SEP-2002 (first entry)

X E Mouse laminin 10 second chain cDNA sequence SEQ ID NO:11.

X W Laminin alpha 5; laminin 10; vulnary; cell growth; differentiation;

X W tissue repair development; laminin; healing; vascular tissue;

X W re-endothelialisation; vascular injury; cell attachment; cell stasis;

X W proliferation; migration; gene; ss.

X S Mus musculus.

X H Key Location/Qualifiers

X T 1..5178

X T /tag= a

X T /partial

X T /product= "laminin 10 second chain"

X T /note= "no start codon given"

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

X T

WO200250111-A2.

27-JUN-2002.

21-DEC-2001; 2001WO-US051035.

21-DEC-2000; 2000US-0257449P.

28-MAR-2001; 2001US-0279282P.

13-NOV-2001; 2001US-00279282.

(BIO-) BIOSTRATUM INC.

Tryggvason K, Doi M, Thyboll J;

WPI; 2002-557650/59.

P-PSDB; ABB81593.

New human laminin-10 proteins, useful for accelerating the healing of vascular tissue, improving the biocompatibility of grafts, or for promoting re-endothelialization at the site of vascular injuries.

Claim 8; Page 145-153; 231pp; English.

The present invention describes human laminin alpha 5. Also described is an isolated laminin 10. Laminin 10 has vulnary activity. Laminins are useful in maintaining cell/tissue phenotype as well as promoting cell growth and differentiation in tissue repair development. Specifically, laminin 10 can be used for accelerating the healing of vascular tissue, improving the biocompatibility of grafts useful for treating such injuries, for promoting re-endothelialization at the site of vascular injuries, and promote cell attachment and subsequent cell stasis, proliferation, differentiation, and/or migration. The present invention encodes a second chain protein of laminin 10, from the present invention

Alignment Scores:

Pred. No.:	358	Length:	5329
Score:	42.00	Matches:	7
Percent Similarity:	71.43%	Conservative:	3
Best Local Similarity:	50.00%	Mismatches:	4
Query Match:	53.18%	Indels:	0
DB:	6	Gaps:	0

US-09-737-297-3 (1-16) x ABO72911 (1-5329)

QY 3 GlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16
|||||
Db 3723 GGCAGACAGCTGGAGTTTATCAAAACTCCGATATTCAGGG 3764

RESULT 11

AAA8899

ID AAA8899 standard; cDNA; 5689 BP.

XX

AC AAA8899;

XX

DT 05-MAR-2001 (first entry)

XX

DE Mouse laminin 2 beta-1 chain cDNA.

XX

KW Laminin 2; mouse; nerve regeneration; angiogenic; cell adhesion;

KW degenerative muscle disorder; muscular dystrophy; cell therapy; ss.

XX

OS Mus musculus.

XX

FH Key Location/Qualifiers

FT CDS 178..5538

FT /tag= a

FT sig_peptide 178..240

FT /tag= b

FT mat_peptide 241..5541

FT /tag= c

XX

PN WO200066730-A2.

XX

XX

PD 09-NOV-2000.

XX

XX

PF 28-APR-2000; 2000WO-US011378.

XX

PR 30-APR-1999; 99US-0131720P.

PR 15-JUN-1999; 99US-0139138P.

PR 12-JUL-1999; 99US-0143289P.

PR 24-SEP-1999; 99US-0155945P.

XX

PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX

PI Yurchenco P;

XX

XX

WPI; 2000-687537/67.

XX

P-PSDB; AAB19799.

XX

XX

PS Claim 4; Page 205-212; 305pp; English.

XX

XX

CC The present sequence is that of cDNA encoding the beta-1 chain (see

CC AAB19799) of mouse laminin 2. Laminin 2 is composed of alpha-2 (400 kDa),

CC beta-1 (100 kDa) and gamma-1 (100 kDa) chains. It is thought to be

CC specifically required for stabilizing myotubes during skeletal muscle

CC development, and for preventing apoptosis. Genetic defects in human

CC laminin 2 structure or expression are associated with a major type of

CC congenital muscular dystrophy. Laminin 2 is also thought to be important

CC in Schwann cell/basal lamina interactions. The invention provides laminin

CC 2 alpha-2, beta-1 and gamma-1 chain polypeptides (see AAB19791-806) and

CC the polynucleotides encoding them (see AAA8891-906), methods for making

CC recombinant laminin 2, cells that express recombinant laminin 2, and

CC methods for using purified laminin 2 for research and therapeutic

CC purposes including peripheral nerve regeneration, treatment of

CC degenerative muscle disorders, angiogenesis regulation, promoting cell

CC attachment and migration, ex vivo cell therapy, improving the take of

CC grafts, improving the biocompatibility of medical devices and preparing

CC improved culture devices and media

XX

SQ Sequence 5689 BP; 1462 A; 1409 C; 1570 G; 1248 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 388 Length: 5689
 Score: 42.00 Matches: 7
 Percent Similarity: 71.43% Conservatives: 3
 Best Local Similarity: 50.00% Mismatches: 4
 Query Match: 53.16% Indels: 0
 DB: 3 Gaps: 0

JS-09-737-297-3 (1-16) x AAA88899 (1-5689)
 2Y 3 GlySerThr***AspValTYrGlnAsnIleGlnTYrAlaGly 16
 Db 4083 GGCAGAACGCTGGATTATCAAAAACTCCGATATTCAGGG 4124

RESULT 12
 AAC83711
 ID AAC83711 standard; cDNA; 5689 BP.
 XX
 AC AAC83711;
 XX
 DT 02-MAR-2001 (first entry)
 XX
 DE Mouse laminin 8 cDNA, SEQ ID NO: 17.
 XX
 KW Mouse; laminin 8; neuroprotective; angiogenic; osteopathic;
 KW antiarteriosclerotic; glycoprotein; mesenchymal tissue injury;
 KW vascular tissue injury; neural injury; angiogenesis regulation; ss.
 XX
 OS Mus musculus.

XX
 FN WO20006732-A2.
 XX
 PD 09-NOV-2000.
 XX
 PF 28-APR-2000; 2000WO-US011543.
 XX
 PR 30-APR-1999; 99US-0131720P.
 PR 21-AUG-1999; 99US-0149738P.
 PR 24-SEP-1999; 99US-0155945P.
 PR 11-FEB-2000; 2000US-0182012P.
 XX
 PA (BIOS-) BIOSTRATUM INC.

XX
 PI Kortessmaa J, Tryggvason K;
 XX
 DR WPI; 2000-687539/67.
 DR P-PSDB; AAB48450.
 XX
 PT Purified laminin 8 protein, useful for research and therapeutic purposes
 PT including peripheral nerve regeneration, treatment of degenerative muscle
 PT disorders, angiogenesis regulation, and ex vivo cell therapy.
 XX
 PS Claim 4; Page 169-176; 245pp; English.

XX
 CC The present sequence encodes a laminin 8 polypeptide chain. Laminins are
 CC a family of heterotrimeric glycoproteins that function via binding
 CC interactions with neighbouring cell receptors and by forming laminin
 CC networks. They are signalling molecules which influence cellular
 CC function. Laminin 8 is useful for treating injuries to tissue of
 CC mesenchymal origin, such as bone, cartilage, tendon, and ligament.
 CC treating injuries to vascular tissue, promoting cell attachment and
 CC migration, ex vivo cell therapy, improving the biocompatibility of
 CC medical devices, and preparing improved cell culture devices and media.
 CC Laminin 8 is also useful for promoting re-endothelialisation at the site
 CC of vascular injuries, improving the take of grafts, improving the
 CC biocompatibility of medical devices, treating neural injuries (neural
 CC regeneration), regulating angiogenesis, and promoting cell attachment and
 CC migration.

XX
 SQ Sequence 5689 BP; 1462 A; 1409 C; 1570 G; 1248 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 388 Length: 5689

Score: 42.00 Matches: 7
 Percent Similarity: 71.43% Conservatives: 3
 Best Local Similarity: 50.00% Mismatches: 4
 Query Match: 53.16% Indels: 0
 DB: 3 Gaps: 0

US-09-737-297-3 (1-16) x AAC83711 (1-5689)
 OY 3 GlySerThr***AspValTYrGlnAsnIleGlnTYrAlaGly 16
 Db 4083 GGCAGAACGCTGGATTATCAAAAACTCCGATATTCAGGG 4124

RESULT 13
 ABQ72910
 ID ABQ72910 standard; cDNA; 5689 BP.
 XX
 AC ABQ72910;
 XX
 DT 19-SEP-2002 (first entry)
 XX
 DE Mouse laminin 10 second chain cDNA sequence SEQ ID NO:9.
 XX
 KW Laminin alpha 5; laminin 10; vulnery; cell growth; differentiation;
 KW tissue repair development; laminin; healing; vascular tissue;
 KW re-endothelialisation; vascular injury; cell attachment; cell stasis;
 KW proliferation; migration; gene; ss.

XX
 OS Mus musculus.
 XX
 FN Key Location/Qualifiers
 FT CDS 178..5538
 FT /*tag= a
 FT /product= "laminin 10 second chain"
 FT sig_peptide 178..240
 FT /*tag= b
 FT mat_peptide 241..5535
 FT /*tag= c
 FT /product= "laminin 10 second chain"
 XX
 FN WO200250111-A2.

XX
 PD 27-JUN-2002.
 XX
 PF 21-DEC-2001; 2001WO-US051035.
 XX
 PR 21-DEC-2000; 2000US-0257449P.
 PR 28-MAR-2001; 2001US-0279282P.
 PR 13-NOV-2001; 2001US-00279282.
 XX
 PA (BIOS-) BIOSTRATUM INC.
 XX
 PI Tryggvason K, Doi M, Thyboll J;
 XX
 DR WPI; 2002-557650/59.
 DR P-PSDB; ABB81592.

XX
 PT New human laminin-10 proteins, useful for accelerating the healing of
 PT vascular tissue, improving the biocompatibility of grafts, or for
 PT promoting re-endothelialization at the site of vascular injuries.
 XX
 PS Claim 8; Page 132-140; 23pp; English.

XX
 CC The present invention describes human laminin alpha 5. Also described is
 CC an isolated laminin 10. Laminin 10 has vulnery activity. Laminins are
 CC useful in maintaining cell/tissue phenotype as well as promoting cell
 CC growth and differentiation in tissue repair development. Specifically,
 CC laminin 10 can be used for accelerating the healing injuries of vascular
 CC tissue, improving the biocompatibility of grafts useful for treating such
 CC injuries, for promoting re-endothelialization at the site of vascular
 CC injuries, and promote cell attachment and subsequent cell stasis,
 CC proliferation, differentiation, and/or migration. The present sequence
 CC encodes a second chain protein of laminin 10, from the present invention

CC for *S. pneumoniae*
XX
SQ Sequence 11384 BP; 3543 A; 2446 C; 2004 G; 3391 T; 0 U; 0 Other;

Alignment Scores:		
Pred. No.:	909	Length: 11384
Score:	42.00	Matches: 7
Percent Similarity:	76.92%	Conservative: 3
Best Local Similarity:	53.85%	Mismatches: 3
Query Match:	53.16%	Indels: 0
DB:	2	Gaps: 0

Y 3 GlySerThr**AspValTyrGlnAsnIleGlnTyrAlaGly 16

Qy	2	GluGlySerThr***AspValTyrGlnAsnIleGlnTyr	14
		: : : : :	
		: : : : :	
Db	8894	GAGGGACTAGTCACGCTATTTACCAAAATATCGACTAT	8856

RESULT 14
AV52178/C

US-09-737-297-3 (1-16) x AAV52178 (1-11384)

AAV52178;

23-OCT-1998 (first entry)

Streptococcus pneumoniae genome fragment

X W Streptococcus pneumoniae: S. pneumoniae: genome: diagnosis: assay:

computer readable medium; vaccine; pharmaceutical composition; ds.

ХИЩНИКИ И ЖИВОТНЫЕ

XX
IN
COUNCIL OF THE EUROPEAN UNION
1999

[illegible]

30-CCI-1997; 97MO-US019588.

31-OCT-1996; 96US-0029960P.

AA (HOMA²) HUMAN GENOME SCI INC.

ROSEN CA, BARASH SC, FANNON M,
DILLON FU, CHOI GH, KUNISCH CA,
Dougherty BA:

WPI: 1998-272225/24.

Computer-readable medium with recorded streptococcus pneumoniae

polynucleotide sequences - useful in diagnostic kits and assays, pharmaceutical compositions and vaccines for streptococcal pneumonia

Claim 1; Page 416-423: 1409pp; English:

The present invention describes a computer readable medium which has

nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) on it, or a representative fragment or a sequence at least 95% identical

to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus

CC pneumoniae: the present invention also describes an isolated nucleic acid molecule encoding a homologue of any of the fragments of the sequence of the present invention.

genome (SEQ ID NO:1 to 351) where the nucleic acid molecule is produced by a process comprising: (a) screening a genomic DNA library using

probe a target sequence defined by any of the sequences in SEQ ID NO: 391, identifying members of the library which contain sequences that

hybridize to the target sequence and isolating the nucleic acid from the members: or (b) isolating mRNA, DNA or cDNA produced from the members; or

organism, amplifying nucleic acid molecules whose nucleotide sequence is homologous to amplification primers derived from the fragment of the

pneumoniae genome to prime the amplification and isolating the amp sequences. The computer readable medium can be used in a computer-

system for identifying fragments of the *S. pneumoniae* genome of commercial importance, or expression modulating fragments of the

pneumoniae genome. Products from the present invention can be used in diagnosis kits and assays, and pharmaceutical compositions and vaccine

S Homo sapiens.
X WO200214500-A2.
X X
X D 21-FEB-2002.
X X

X F 15-AUG-2001; 2001WO-US025840.
X X

X R 16-AUG-2000; 2000US-0226326P.
X X

X A (CHIR) CHIRON CORP.
X A (HYSE-) HYSEQ INC.

X Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F,
X Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;
X WPI; 2002-241905/29.

X New nucleic acid for producing a polypeptide, detecting differentially
X expressed genes correlated with a cancerous state of a mammalian cell,
X and inhibiting tumor growth.

X Claim 1; SEQ ID NO 2064; 883pp + Sequence Listing; English.

X The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
X with cytostatic activity. The polynucleotide is used to produce a
X polypeptide to detect differentially expressed genes correlated with a
X cancerous state of a mammalian cell and to inhibit tumor growth. The
X polynucleotide is used as a probe in mapping and tissue profiling. The
X encoded polypeptide and antibodies to the polypeptide can also be used
X for therapeutic and diagnostic purposes. The polynucleotide is useful for
X gene therapy. Note: The sequence data for this patent did not form part
X of the printed specification, but was obtained in electronic format
X directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

X Sequence 125 BP; 38 A; 27 C; 16 G; 44 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 5.79 Length: 125
Score: 41.00 Matches: 8
Percent Similarity: 64.29% Conservative: 1
Best Local Similarity: 57.14% Mismatches: 5
Query Match: 51.90% Indels: 0
DB: 6 Gaps: 0

US-09-737-297-3 (1-16) x ABN62097 (1-125)

Qy 2 GlucylserThr***AspValTyrGlnAsnIleGIntyAla 15
Db 44 GAGGGCTCAGTCATAGATGCTATGATGAGATATATGCGAGCT 3

RESULT 17

ABV46101

ID ABV46101 standard; cDNA; 441 BP.

XX AC ABV46101;

XX 16-SEP-2002 (first entry)

DE Human prostate expression marker cDNA 46092.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.

XX Homo sapiens.

XX WO200160860-A2.

XX 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US005171.

XX 17-FEB-2000; 2000US-0183319P.

XX

PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211324P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JB;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer.

XX Claim 1; Page 9105; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX specification or its complement. (I) is useful for: (a) assessing whether
XX a patient is afflicted with prostate cancer; (b) monitoring the
XX progression of prostate cancer in a patient; (c) assessing the efficacy
XX of a test compound to inhibit prostate cancer in a patient; (d) assessing
XX the efficacy of a therapy for inhibiting prostate cancer in a patient;
XX (e) selecting a composition for inhibiting prostate cancer in a patient;
XX (f) assessing the prostate cell carcinogenic potential of a compound; (g)
XX determining whether prostate cancer has metastasized in a patient; (h)
XX assessing the aggressiveness or indolence of prostate cancer in a patient
XX ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

XX Sequence 441 BP; 154 A; 84 C; 74 G; 129 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 27.3 Length: 441
Score: 41.00 Matches: 8
Percent Similarity: 56.25% Conservative: 1
Best Local Similarity: 50.00% Mismatches: 7
Query Match: 51.90% Indels: 0
DB: 5 Gaps: 0

US-09-737-297-3 (1-16) x ABV46101 (1-441)

Qy 1 AlaGluGlySerThr***AspValTyrGlnAsnIleGIntyAlaGly 16

Db 349 GCTCCCAAAAGACATGGATTGTACCAAAACACACACTACAAAGGC 396

RESULT 18

ABN76334

ID ABN76334 standard; cDNA; 445 BP.

XX AC ABN76334;

XX 08-JUL-2002 (first entry)

DE Human ORF1281 cDNA, SEQ ID NO:2561.

XX Human; ORF; open reading frame; ORF; drug screening; diagnosis;
XX disease monitoring; cytokine; cell proliferation; cell differentiation;
XX immune modulation; haematopoiesis regulation; tissue growth;
XX angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
XX thrombolytic; tumour inhibition; bodily characteristic; fertility;
XX behaviour; cancer; proliferative disorder; neurological disorder;
XX cardiovascular disease; immune system disorder; organ transplantation;
XX tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
XX hypothyroidism; cholesterol ester storage disease; infection; vulnery;
XX vasotropic; antipsoriatic; antidiabetic; cytostatic; neurotropic;
XX neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
XX cardiac; hypotensive; antithyroid; antiinflammatory; immunomodulator;
XX dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.

XX Homo sapiens.

N WC200190366-A2.
X 29-NOV-2001.
X 24-MAY-2001; 2001WO-US017076.
F 24-MAY-2000; 2000US-0206690P.
R (CURA-) CURAGEN CORP.
X Leach MD, Shinkets RA;
X WPI; 2002-106200/14.
X P-PSDB; ABP32308.
X Novel human polypeptides and polynucleotides useful for diagnosing,
X preventing and treating cardiovascular disease, neurodegenerative,
X hyperproliferative disorders and disorders related to organ
X transplantation.
X Claim 1; Page 901; 2508pp; English.
X Sequences ABP31028-ABP3561 represent 4534 novel human proteins
X designated ORF (open reading frame) 1-4534, and sequences ABN75054-
X ABN7587 represent cDNAs encoding them. The invention also encompasses
X polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
X referred to as ORFX) proteins, polynucleotides at least 85% identical to
X the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
X polynucleotides, the recombinant production of ORFX proteins, antibodies
X specific for ORFX proteins, methods of detecting ORFX polynucleotides and
X polypeptides, methods of screening for modulators of ORFX expression or
X activity, and methods of screening individuals for a predisposition to an
X ORFX-associated disorder. The ORFX proteins of the invention have a wide
X range of biological activities, such as cytokine, cell proliferation,
X cell differentiation, immune modulation, haematopoiesis regulation,
X tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
X chemokinetic activity, haemostatic activity, thrombolytic activity,
X receptor/ligand, antiinflammatory activity, tumour inhibition activity,
X and antiinfective activity, and may also be involved in the determination
X of bodily characteristics, fertility and behaviour. ORFX proteins,
X nucleic acids and antibodies may be used in the treatment of cancers,
X other proliferative disorders such as psoriasis and benign tumours,
X neurological disorders such as epilepsy and Alzheimer's disease,
X cardiovascular diseases, immune system disorders, disorders related to
X organ transplantation, disorders of tissue growth and regeneration,
X diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
X storage disease, and infectious diseases caused by viral, bacterial,
X fungal and other pathogens. ORFX nucleic acids may also be used as a
X source of primers and probes, in the detection of ORFX genomic sequences
X or transcripts, in the identification and cloning of homologous
X sequences, in genetic diagnosis, and in forensic biology. The ORFX
X nucleic acids may additionally be used to produce transgenic animals
X which may be useful for studying the function and/or activity of ORFX
X protein, and in drug screening. The ORFX proteins may also be used as
X immunogens to generate specific antibodies, which are useful in the
X diagnosis, treatment and monitoring of ORFX-associated diseases
X
X SQ Sequence 445 BP; 91 A; 138 C; 74 G; 142 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 27.6 Length: 445
Score: 41.00 Matches: 7
Percent Similarity: 84.62% Conservative: 4
Best Local Similarity: 53.85% Mismatches: 2
Query Match: 51.90% Indels: 0
DB: 6 Gaps: 0
US-09-737-297-3 (1-16) x ABN76334 (1-445)
Qy 3 GlySerThr***AspValTrpGlnAsnIleGlnTrpAla 15
Db 273 GGCAGCTGGGTGATCTCTATTAATTCGCACTGGACT 311

RESULT 19
AAQ36563
ID AAQ36563 standard; DNA; 675 BP.
XX AC AAQ36563;
XX DT 25-MAR-2003 (revised)
XX DT 15-JUN-1993 (first entry)
XX DE Endo-xylanase gene from Streptomyces.
XX XlnC; xylosidic linkages; pulp; lignocellulose; brightening; viscosity;
XX swelling; ss.
XX OS Streptomyces lividans.
XX FH Key Location/Qualifiers
XX CDS 1..603 /*tag= a
XX FT sig_peptide 1..27 /*tag= b
XX FT mat_peptide 28..603 /*tag= c
XX W09303155-A1.
XX 18-FEB-1993.
XX 10-AUG-1992; 92WO-CA000349.
XX 08-AUG-1991; 91CA-02048322.
XX (FRAP-) FRAPPIER ARMAND INST.
XX Kluepfel D, Morosoli R, Shareck F;
XX WPI; 1993-076512/09.
XX P-PSDB; AAR32263.
XX New high-activity endo-xylanase from Streptomyces - used for treating
XX lignocellulose, pref. pulp for delignification, brightening and viscosity
XX improvement.
XX Claim 32; Page 8; 30pp; English.
XX Chromosomal DNA was extd. from S. lividans 66 (strain 1326) and
XX restriction fragments sepd. The strain S. lividans strain 1326 and
XX mutated using N-methyl-N'-nitro-N-nitrosoguanidine and a double mutant
XX beta-1,4-D-glucan glucanohydrolase (endocellulase)-negative and xylanase
XX -negative was selected. Prototyping and transformation of the double
XX mutant was carried out using the DNA fragments and the multicopy vector
XX pIJ702 to obtain the xylanase C- harbouring multicopy clone xlnC. The
XX xylanase enzyme produced by this clone has a higher activity than known
XX xylanases and can be used for hydrolysing beta-1,4-D-xylosidic linkages
XX in pulp. The treatment of lignocellulose material with the xylanase
XX results in delignification, brightening and viscosity improvement.
XX Further, such treatment may provide more relaxed fibres resulting in an
XX improved performance or a subsequent treatment, such as swelling,
XX beating, drainage or chemical bleaching of the pulp, with an overall
XX reduction in energy and chemicals used. (Updated on 25-MAR-2003 to
XX correct PN field.)
XX SQ Sequence 675 BP; 139 A; 233 C; 202 G; 101 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 46 Length: 675
Score: 41.00 Matches: 6
Percent Similarity: 71.43% Conservative: 4
Best Local Similarity: 42.86% Mismatches: 4
Query Match: 51.90% Indels: 0
DB: 2 Gaps: 0
US-09-737-297-3 (1-16) x AAQ36563 (1-675)

2y 1 AlaGluGlySerThr***AspValTyTGlnAsnIleGlnTyr 14
2b 346 AGCGACGGAGCCCTACGACACTTACCGACGCCGCTAC 387

RESULT 20
AD53545/c
ID AAD53545 standard; DNA; 799 BP.

XX AAD53545;
XX 28-MAY-2003 (first entry)

XX Woody plant hybrid aspen PttIAA4 DNA.
XX Aux/IAA gene; PttIAA; auxin sensitivity; wood quality; plant fiber; gene;
XX ds.

XX Populus tremula.
XX Populus tremuloides.

XX Key Location/Qualifiers
XX CDS 21..551
XX /*tag= a
XX /product= "PttIAA4 protein"
XX /transl_except= (pos:258..269, aa:Asn-Ser)

XX WO200292823-A1.
XX 21-NOV-2002.

XX 14-MAY-2002; 2002WO-S0000912.
XX 14-MAY-2001; 2001SE-00001694.

XX (SWEET-) SWEETREE GENOMICS AB.
XX Bhalarao R, Schrader J, Moyle R, Sandberg G;
XX WPI; 2003-120690/11.
XX P-PSDB; AAE35121.

XX New PttIAA genes 1-8, useful for regulating auxin sensitivity in a plant,
XX for altering the wood quality or quantity in a woody plant, and for
XX modifying and controlling the synthesis of plant fibers.

XX Claim 1; Col 40; 31pp; English.
XX The present invention relates to polynucleotides and amino acid sequences
XX of Aux/IAA gene family, designated as PttIAA1-8. PttIAA sequences of the
XX invention are useful for regulating auxin sensitivity in a plant, for
XX altering the wood quality or quantity in a woody plant and for modifying
XX and controlling the synthesis of plant fibers. The present sequence is
XX PttIAA4 DNA from the woody plant hybrid aspen (Populus tremula x Populus
XX tremuloides)

XX Sequence 799 BP; 239 A; 165 C; 172 G; 223 T; 0 U; 0 Other;
XX Alignment Scores:
XX Pred. No.: 56.7 Length: 799
XX Score: 41.00 Matches: 7
XX Percent Similarity: 81.82% Conservative: 2
XX Best Local Similarity: 63.64% Mismatches: 2
XX Query Match: 51.90% Indels: 0
XX DB: 7 Gaps: 0

US-09-737-297-3 (1-16) x AAD53545 (1-799)
Qy 4 SerThr***AspValTyTGlnAsnIleGlnTyr 14
Db 444 TCACACATGACATGCCGAGATGTCATAT 412

RESULT 21

AAS77451
ID AAS77451 standard; CDNA; 1078 BP.

XX AAS77451;
XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #13255.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ABG13264.

XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.

XX Claim 1; SEQ ID NO 13255; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridization probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
XX coding sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pcr_sequences

XX Sequence 1078 BP; 445 A; 237 C; 179 G; 216 T; 0 U; 0 Other;
XX Alignment Scores:
XX Pred. No.: 81.9 Length: 1078
XX Score: 41.00 Matches: 7
XX Percent Similarity: 66.67% Conservative: 3
XX Best Local Similarity: 46.67% Mismatches: 5
XX Query Match: 51.90% Indels: 0
XX DB: 5 Gaps: 0

US-09-737-297-3 (1-16) x AAS77451 (1-1078)
Qy 2 GluGlySerThr***AspValTyTGlnAsnIleGlnTyrAlaGly 16

b 136 GATGATTACACAGCGGAATTTACACAGAGATAAAATACCTGGGA 180

RESULT 22

AS5336/c

D AAS5336 standard; DNA; 1230 BP.

X C AAS5336;

X T 13-FEB-2002 (first entry)

X E Haemophilus influenzae DNA for cellular proliferation protein #318.
X W Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
X W antibacterial; drug design.
X S Haemophilus influenzae.

X N WO200170955-A2.

X D 27-SEP-2001.

X F 21-MAR-2001; 2001WO-US009180.

X R 21-MAR-2000; 2000US-0191078P.

X R 23-MAY-2000; 2000US-0206848P.

X R 26-MAY-2000; 2000US-0207272P.

X R 23-OCT-2000; 2000US-0242578P.

X R 27-NOV-2000; 2000US-0253625P.

X R 22-DEC-2000; 2000US-0257931P.

X R 16-FEB-2001; 2001US-0269308P.

X A (ELIT-) ELITRA PHARM INC.

X I Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

X I Yamamoto RT, Xu HH;

X R WPI; 2001-611495/70.

X R P-PSDB; AAU35677.

X T New polynucleotides for the identification and development of

X T antibiotics, comprise sequences of antisense nucleic acids.

X X Claim 27; SEQ ID NO 7173; 511pp; English.

X X The invention relates to antisense inhibitors of genes essential to
X X prokaryotic cellular proliferation, their use in identifying the genes,
X X themselves and the encoded proteins. The prokaryotes used are Escherichia
X X coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
X X Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
X X useful for the identification of potential new targets for antibiotic
X X development. The antisense nucleic acids can also be used to identify
X X proteins used in proliferation, to express these proteins, and to obtain
X X antibodies capable of binding to the expressed proteins. The proteins can
X X be used to screen compounds in rational drug discovery programmes. The
X X antisense nucleic acid sequence is also useful to screen for homologous
X X of organisms. The present sequence encodes an essential prokaryotic
X X cellular proliferation protein. Note: The sequence data for this patent
X X did not form part of the printed specification, but was obtained in
X X electronic format directly from WIPO at
X X ftp.wipo.int/pub/published_pct_sequences

X X Sequence 1230 BP; 387 A; 220 C; 277 G; 346 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 96.3 Length: 1230
Score: 41.00 Matches: 7
Percent Similarity: 71.43% Conservative: 3
Best Local Similarity: 50.00% Mismatches: 4
Query Match: 51.90% Indels: 0
Gaps: 0

US-09-737-297-3 (1-16) x AAS5336 (1-1230)

QY 1 AlaGluGlySerThr***AspValTyrGlnAsnIleGlnTyr 14

Db 565 GCAACGGGTCATGGCACACGTTTTCAGACGTCGGCTAT 524

RESULT 23

AA573157

ID AAS73157 standard; CDNA; 1626 BP.

XX AC AAS73157;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #8961.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-PSDB; ABG08970.

XX PT New isolated polynucleotide and encoded polypeptides, useful in

XX diagnostics, forensics, gene mapping, identification of mutations

XX responsible for genetic disorders or other traits and to assess

XX biodiversity.

XX PS Claim 1; SEQ ID NO 8961; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
X X sequences. (I) is useful as hybridisation probes, polymerase chain
X X reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
X X and in recombinant production of (II). The polynucleotides are also used
X X in diagnostics as expressed sequence tags for identifying expressed
X X genes. (I) is useful in gene therapy techniques to restore normal
X X activity of (II) or to treat disease states involving (II). (II) is
X X useful for generating antibodies against it, detecting or quantitating a
X X polypeptide in tissue, as molecular weight markers and as a food
X X supplement. (II) and its binding partners are useful in medical disorders
X X of sites expressing (II). (I) and (II) are useful for treating disorders
X X involving aberrant protein expression or biological activity. The
X X polypeptide and polynucleotide sequences have applications in
X X diagnostics, forensics, gene mapping, identification of mutations
X X responsible for genetic disorders or other traits to assess biodiversity
X X and to produce other types of data and products dependent on DNA and
X X amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
X X coding sequences of the invention. Note: The sequence data for this
X X patent did not appear in the printed specification, but was obtained in
X X electronic format directly from WIPO at
X X ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1626 BP; 686 A; 352 C; 265 G; 323 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 136 Length: 1626
Score: 41.00 Matches: 7
Percent Similarity: 66.67% Conservative: 3
Best Local Similarity: 46.67% Mismatches: 5

S-09-737-297-3 (1-16) x ABK98515 (1-2139)
Y 2 GluGlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 14
b 1001 GACAAACCCACCGCGGACCTCTACACGAGCAATTAATACCTGGGA 1039
RESULT 26
AS69826
D AAS69826 standard; cDNA; 4683 BP.
X C AAS69826;
X T 13-FEB-2002 (first entry)
X E DNA encoding novel human diagnostic protein #5630.
X W Human; chromosome mapping; gene mapping; gene therapy; forensic;
X W food supplement; medical imaging; diagnostic; genetic disorder; as.
X S Homo sapiens.
X N WO200175067-A2.
X D 11-OCT-2001.
X F 30-MAR-2001; 2001WO-US008631.
X R 31-MAR-2000; 2000US-00540217.
X R 23-AUG-2000; 2000US-00649167.
X A (HYSE-) HYSEQ INC.
X I Drmanac RT, Liu C, Tang YT;
X K WPI; 2001-639362/73.
X R P-PSDB; ABG05639.
X X New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
PS Claim 1; SEQ ID NO 5630; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

Sequence 4683 BP; 1646 A; 901 C; 928 G; 1208 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 498 Length: 4683
Score: 41.00 Matches: 7
Percent Similarity: 66.67% Conservative: 3

Best Local Similarity: 46.67% Mismatches: 5
Query Match: 51.90% Indels: 0
Db: 5 Gaps: 0
US-09-737-297-3 (1-16) x AAS69826 (1-4683)
QY 2 GluGlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16
Db 136 GATGGATTTCACGCGCAATTCACAGAGAAATAAATACCTGGGA 180
RESULT 27
ABL23219
ID ABL23219 standard; DNA; 8301 BP.
XX AC ABL23219;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 21130.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX Claim 1; SEQ ID NO 21130; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABU01840-ABU16175) and the encoded proteins (ABBS57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 8301 BP; 2301 A; 2095 C; 2104 G; 1801 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 1.01e+03 Length: 8301
Score: 41.00 Matches: 8
Percent Similarity: 60.00% Conservative: 1
Best Local Similarity: 53.33% Mismatches: 6
Query Match: 51.90% Indels: 0
Db: 4 Gaps: 0
US-09-737-297-3 (1-16) x ABL23219 (1-8301)
QY 2 GluGlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16
Db 1595 GAGGGTTTCCACCGGATGTAAAGCAGATCTAAACTACTGTGGG 1639

DC00771
D ADC00771 standard; DNA; 43450 BP.
X
C ADC00771;
X
T 04-DEC-2003 (first entry)
X
E Enterohaemorrhagic E. coli O157:H7-specific nucleic acid SEQ ID NO: 816.
X
W ds; gene; enterohaemorrhagic; anti-bacterial.
X
S Escherichia coli; O157:H7.
X
N JP2002355074-A.
X
D 10-DEC-2002.
X
F 24-JAN-2002; 2002JP-00015959.
X
R 24-JAN-2001; 2001JP-00112010.
X
A (UYTS-) UNIV TSUKUBA.
X
R WPI; 2003-451640/43.
X
T Enterohaemorrhagic Escherichia coli O157:H7-specific nucleic acid molecule
X
X and a polypeptide and its use, a polypeptide, a vector and a host cell.
X
PS Claim 2; SEQ ID NO 816; 2067pp; Japanese.
X
C The invention relates to a novel enterohaemorrhagic Escherichia coli
X
C O157:H7-specific nucleic acid molecule. A polynucleotide of the invention
X
C has anti-bacterial activity. The polypeptide can be used in detection
X
C and/or treatment of O157:H7 infection. The nucleotide sequence of the
X
C genome of Enterohaemorrhagic E. coli O157:H7 was determined. The present
X
C sequence represents an E. coli O157:H7-specific nucleic acid of the
X
C invention.
X
SQ Sequence 43450 BP; 13119 A; 9496 C; 8282 G; 12553 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 7.71e+03 Length: 43450
Score: 41.00 Matches: 7
Percent Similarity: 62.50% Conservative: 3
Best Local Similarity: 43.75% Mismatches: 6
Query Match: 51.90% Indels: 0
DB: 9 Gaps: 0

US-09-737-297-3 (1-16) x ADC00771 (1-43450)

QY 1 AlaGluglySerThr***AspValTyGlnAsnIleGlnTyAlaGly 16
Db 30234 TCGTCAGGTTCCACAGATACATTTATCAGACTCTGCAATATCCTGGC 30281

RESULT 31
ACD19247
ID ACD19247 standard; DNA; 45325 BP.
X
X ACD19247;
X
X 27-OCT-2003 (revised)
X
X 21-AUG-2003 (first entry)
X
X E. coli O157 locus of Enterocyte Effacement, LEE.
X
X OZID; ds; acute haemorrhagic colitis; haemolytic uraemic syndrome;
X
X food poisoning; locus of enterocyte effacement; LEE.
X
X Escherichia coli; strain O157:H7.
X
OS US2003023075-A1.
X
X US2003023075-A1.
X
X 30-JAN-2003.
X
PD

XX
PF
XX
PR 01-APR-2002; 2002US-00114170.
XX
PR 04-DEC-1998; 98US-0110955P.
XX
PR 03-DEC-1999; 99US-00453702.
XX
PA (BLAT/) BLATTNER F R.
XX
PA (BURL/) BURLAND V D.
XX
PA (PERN/) PERNA N T.
XX
PA (PLUN/) PLUNKETT G.
XX
PA (WELC/) WELCH R.

Blattner FR, Burland VD, Perna NT, Plunkett G, Welch R;
WPI; 2003-479497/45.

New DNA sequences from Escherichia coli strain O157:H7, useful for
detecting E. coli O157:H7 in a sample, or in designing diagnostic probes
which can be used to distinguish strain O157:H7 from strain K12 using
molecular techniques.

Claim 9; SEQ ID NO 261; 33pp; English.

The invention relates to an isolated DNA molecule comprising an E. coli
strain O157:H7 sequence selected from a clonal cluster, a locus of
a urease gene cluster, a RTX toxin-like gene cluster, a locus of
enterocyte effacement and 2 genes from its associated lymphocytic phage
933W (a putative serine/threonine kinase and a tail fibre gene). E. coli
O157:H7 can cause food poisoning, specifically acute haemorrhagic colitis
(which can develop into haemolytic uraemic syndrome). Also included are
an isolated DNA molecule comprising a nucleotide sequence identical to at
least 25 contiguous nucleotides contained in DNA sequences selected from
ACD18988-ACD19242 (being 255 E. coli O157 DNA sequences which are not
found in E. coli K12), a recombinant DNA construction comprising the DNA
above and a method for detecting E. coli O157:H7 (ATCC 43895) in a sample
(or distinguishing between O157 and K12) using a probe derived from one
of the 255 sequences. The DNA sequences are useful in detecting E. coli
O157:H7 in a sample, for the early diagnosis of humans and livestock
infected with O157:H7, and in designing diagnostic probes which can be
used to distinguish strain O157:H7 from strain K12 using molecular
techniques. The present sequence is the locus of enterocyte effacement,
LEE. Note: The sequence data for this patent did not form part of the
printed specification, but was obtained in electronic format directly
from the USPTO at seqdata.uspto.gov/sequence.html?DocID=20030023075
(Updated on 27-OCT-2003 to standardise OS field)

SQ Sequence 45325 BP; 13582 A; 9936 C; 8732 G; 13075 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8.12e+03 Length: 45325
Score: 41.00 Matches: 7
Percent Similarity: 62.50% Conservative: 3
Best Local Similarity: 43.75% Mismatches: 6
Query Match: 51.90% Indels: 0
DB: 8 Gaps: 0

US-09-737-297-3 (1-16) x ACD19247 (1-45325)

QY 1 AlaGluglySerThr***AspValTyGlnAsnIleGlnTyAlaGly 16
Db 31121 TCGTCAGGTTCCACAGATACATTTATCAGACTCTGCAATATCCTGGC 31168

RESULT 32
AAT42063_17
Continuation (18 of 19) of AAT42063 from base 1700001 (Haemophilus influenzae complete g
WP Sequence split into 19 fragments LOCUS AAT42063 Accession Aat42063
WP Fragment Name Begin End
WP AAT42063_00 1 110000
WP AAT42063_01 100001 210000
WP AAT42063_02 200001 310000
WP AAT42063_03 300001 410000
WP AAT42063_04 400001 510000
WP AAT42063_05 500001 610000

VP AAT42063_06 600001 710000
 VP AAT42063_07 700001 810000
 VP AAT42063_08 800001 910000
 VP AAT42063_09 900001 1010000
 VP AAT42063_10 1000001 1110000
 VP AAT42063_11 1100001 1210000
 VP AAT42063_12 1200001 1310000
 VP AAT42063_13 1300001 1410000
 VP AAT42063_14 1400001 1510000
 VP AAT42063_15 1500001 1610000
 VP AAT42063_16 1600001 1710000
 VP AAT42063_17 1700001 1810000
 VP AAT42063_18 1800001 1830121

Alignment Scores:
 Pred. No.: 2,42e+04 Length: 110000
 Score: 41.00 Matches: 7
 Percent Similarity: 71.43% Conservative: 3
 Best Local Similarity: 50.00% Mismatches: 4
 Query Match: 51.90% Indels: 0
 DB: 2 Gaps: 0

US-09-737-297-3 (1-16) x AAT42063_17 (1-110000)

OY 1 AlaGluGlySerThr***AspValtyrGlnAsnIleGlnTyr 14
 DB 28229 GCAACGGGTGATTGGCACAGTTTTCGACAGTCGCTAT 28270
 RESULT 33
 ABX46781
 ID ABX46781 standard; cDNA; 415 BP.
 XX
 AC ABX46781;
 XX

DT 21-FEB-2003 (first entry)

DE Bovine EST associated with lactation/muscle/fat deposition #11946.

KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
 KW muscle deposition; fat deposition; genome mapping; gene identification;
 KW gene analysis; cattle breeding.

XX Bos Taurus.

OS US2002137139-A1.

XX 26-SEP-2002.

PF 24-SEP-2001; 2001US-00960352.

PR 12-JAN-1999; 99US-0115707P.

PR 11-JAN-2000; 2000US-00480902.

XX (BYAT/) BYATT J C.

PA (MATH/) MATHIALAGAN N.

PA (TAON/) TAO N.

PA (WARR/) WARREN W C.

XX Byatt JC, Mathialagan N, Tao N, Warren WC;

PI WPI; 2003-110599/10.

XX New nucleic acid associated with lactation, and muscle and fat

PT deposition, useful for genome mapping, gene identification and analysis,

PT cattle breeding, or for genetically improving cattle.

XX Claim 2; SEQ ID NO 11946; 245pp; English.

PS The invention relates to a purified nucleic acid molecule associated with

CC lactation or muscle and fat deposition (designated LMFD), derived from

CC cattle, and the LMFD nucleic acid can specifically hybridise to a second

CC nucleic acid molecule comprising any of 15112 nucleotide sequences,

CC appearing as ABX34836-ABX49947, or complements of them. Also included are

CC (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
 CC acid linked to a promoter and a 3' non-translated sequence that
 CC functions in the cell to cause termination of transcription and addition
 CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
 CC (2) determining a level or pattern of a molecule in a bovine cell or
 CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
 CC of the 15112 nucleic acid sequences or its complement or fragment) with a
 CC complementary nucleic acid molecule obtained from the bovine cell or
 CC tissue, where hybridisation between the marker nucleic acid and the
 CC complementary nucleic acid permits the detection of the molecule; and (b)
 CC detecting the level or pattern of the complementary nucleic acid, where
 CC the detection of the complementary nucleic acid is predictive of the
 CC level or pattern of the molecule. The LMFD nucleic acid is used for
 CC determining a level or pattern of a molecule in a bovine cell or tissue.
 CC It is useful for genome mapping, gene identification and analysis, cattle
 CC breeding, preparation of constructs for use in cattle gene expression, or
 CC for genetically improving cattle. The present sequence is one of the
 CC 15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
 CC present sequence was not shown in the specification but was obtained in
 CC electronic format from the USPTO web site:
 CC seqdata.uspto.gov/sequence.html?DocID=20020137139
 XX

SQ Sequence 415 BP; 150 A; 77 C; 87 G; 101 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 41.4 Length: 415
 Score: 40.00 Matches: 6
 Percent Similarity: 71.43% Conservative: 4
 Best Local Similarity: 42.86% Mismatches: 4
 Query Match: 50.63% Indels: 0
 DB: 7 Gaps: 0

US-09-737-297-3 (1-16) x ABX46781 (1-415)

OY 2 GluGlySerThr***AspValtyrGlnAsnIleGlnTyrAla 15
 DB 61 GAGAAACACCGTTGACATCTACCCCAATGTTTATTACAGC 102

RESULT 34

AAH08199/c

ID AAH08199 standard; cDNA; 517 BP.

XX AC AAH08199;

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:5034.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EF1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 03-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-

PT length cDNAs defined in the specification, and for the detection and/or

T diagnosis of the abnormality of the proteins encoded by the full-length
X cDNAs.

X Claim 1; SEQ ID NO 5034; 2537bp + Sequence Listing; English.

X The present invention describes primer sets for synthesizing 5602 full-
C length cDNAs defined in the specification. Where a primer set comprises:
C (a) an oligo-dr primer and an oligonucleotide complementary to the
C complementary strand of a polynucleotide which comprises one of the 5602
C nucleotide sequences defined in the specification, where the
C oligonucleotide comprises at least 15 nucleotides; or (b) a combination
C of an oligonucleotide comprising a sequence complementary to the
C complementary strand of a polynucleotide which comprises a 5'-end
C sequence and an oligonucleotide comprising a sequence complementary to a
C polynucleotide which comprises a 3'-end sequence, where the
C oligonucleotide comprises at least 15 nucleotides and the combination of
C the 5'-end sequence/3'-end sequence is selected from those defined in the
C specification. The primer sets can be used in antisense therapy and in
C gene therapy. The primers are useful for synthesizing polynucleotides,
C particularly full-length cDNAs. The primers are also useful for the
C detection and/or diagnosis of the abnormality of the proteins encoded by
C the full-length cDNAs. The primers allow obtaining of the full-length
C cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
C AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
C represent human amino acid sequences; and AAH13629 to AAH13632 represent
C oligonucleotides, all of which are used in the exemplification of the
C present invention

X Sequence 517 BP; 120 A; 94 C; 92 G; 208 T; 0 U; 3 Other;

Alignment Scores:
Pred. No.: 54.2 Length: 517
Score: 40.00 Matches: 6
Percent Similarity: 90.91% Conservative: 4
Best Local Similarity: 54.55% Mismatches: 1
Query Match: 50.63% Indels: 0
DB: 4 Gaps: 0

JS-09-737-297-3 (1-16) x AAH08199 (1-517)

DY 5 Thr***AspValtyrGlnAsnIleGlnTyrAla 15

DB 35 ACTGAAGATATTATAGATCACTTCATATGCC 3

RESULT 35

AAH01958
ID AAH01958 standard; cDNA; 1104 BP.

XX AC AAH01958;

XX DT 26-MAR-2001 (first entry)

XX DE Murine TH2AF1 isotype #3 encoding cDNA, from TGS mouse.

XX KW Mouse; TH2AF1 isotype; transgenic mouse; TGS; interleukin 9; IL-9;
XX inflammatory response; type 2 helper T-cell; TH; eczema; therapy; atopy;
XX allergy; asthma; rhinitis; urticaria; bronchial hyperresponsiveness;
XX inflammatory bowel disease; IBD; antiasthmatic; antiinflammatory;
XX immunosuppressant; ss.

XX OS Mus musculus.

XX FH Key Location/Qualifiers

XX FT CDS 67..1108

XX FT /tag= a

XX FT /product= "Mouse TH2AF1 isotype #3"

XX PN WO2000066708-A2.

XX PD 09-NOV-2000.

XX PF 01-MAY-2000; 2000WO-US011712.

XX

PR 01-MAY-1999; 99US-0132138P.

XX PA (MAGA-) MAGAININ PHARM INC.

XX PI Louahed J, Dong Q, Levitt RC, Maloy WL, Zhou Y, Nicolaides NC;

XX DR WPI; 2000-656494/63.

XX DR P-PSDB; AAY71906.

XX PT New TH2AF1 genes expressed in association with an inflammatory response
XX in airways mediated by type 2 helper T cells are used to treat asthma, an
XX asthma-related disorder and inflammatory bowel disease.

XX PS Claim 3; Page 79-81; 94pp; English.

XX CC The present sequence is a murine TH2AF1 isotype #3 encoding cDNA obtained
XX from small intestine of TGS mouse. Lungs of transgenic mouse (TGS) which
XX over-expresses IL-9 (interleukin-9) is used to isolate IL-9 induced
XX genes. TH2AF1 gene is expressed in association with an inflammatory
XX response in the airways mediated by type 2 helper T-cells (TH). The
XX TH2AF1 genes are selectively upregulated by IL-9 and are part of the IL-9
XX signalling pathway. Down regulation of TH2AF1 is used in the diagnosis,
XX prevention or treatment of atopic allergy including asthma, bronchial
XX hyperresponsiveness, rhinitis, urticaria, allergic inflammatory bowel
XX disease (IBD) and various forms of eczema

XX SQ Sequence 1104 BP; 283 A; 270 C; 298 G; 253 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 138 Length: 1104
Score: 40.00 Matches: 8
Percent Similarity: 71.43% Conservative: 2
Best Local Similarity: 57.14% Mismatches: 4
Query Match: 50.63% Indels: 0
DB: 3 Gaps: 0

US-09-737-297-3 (1-16) x AAH01958 (1-1104)

OY 1 AlaGluGlySerThr***AspValtyrGlnAsnIleGlnTyr 14

DB 442 GCAGAGGGTCCCAAGTGTGATGACTACAGAACCCCTGGCTAC 483

RESULT 36

AAH53136
ID AAH53136 standard; DNA; 1371 BP.

XX AC AAH53136;

XX DT 13-FEB-2002 (first entry)

XX DE Enterococcus faecalis DNA for cellular proliferation protein #564.

XX KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
XX antibacterial; drug design.

XX OS Enterococcus faecalis.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US009180.

XX PR 21-MAR-2000; 2000US-0191078P.

XX PR 23-MAY-2000; 2000US-0206848P.

XX PR 26-MAY-2000; 2000US-0207727P.

XX PR 23-OCT-2000; 2000US-0242578P.

XX PR 27-NOV-2000; 2000US-0253625P.

XX PR 22-DEC-2000; 2000US-0257931P.

XX PF 16-FEB-2001; 2001US-0269308P.

XX PA (ELIT-) ELITRA PHARM INC.

XX

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX WPI; 2001-611495/70.
XX P-PSDB; AAU35277.
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids.
XX Claim 27; SEQ ID NO 6773; 511pp; English.
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the genes,
XX their use in the discovery of novel antibiotics, the essential genes
XX themselves and the encoded proteins. The prokaryotes used are Escherichia
XX coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
XX Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
XX useful for the identification of potential new targets for antibiotic
XX development. The antisense nucleic acids can also be used to identify
XX proteins used in proliferation, to express these proteins, and to obtain
XX antibodies capable of binding to the expressed proteins. The proteins can
XX be used to screen compounds in rational drug discovery programmes. The
XX antisense nucleic acid sequence is also useful to screen for homologous
XX of organisms. The present sequence encodes an essential prokaryotic
XX cellular proliferation protein. Note: The sequence data for this patent
XX did not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1371 BP; 436 A; 236 C; 296 G; 403 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 180 Length: 1371
Score: 40.00 Matches: 7
Percent Similarity: 60.00% Conservative: 2
Best Local Similarity: 46.67% Mismatches: 6
Query Match: 50.63% Indels: 0
DB: 4 Gaps: 0
US-09-737-297-3 (1-16) x AAS53136 (1-1371)
QY 2 GlucySerThr***AspValTyGlnAsnIleGlnTyAlaGly 16
DB 763 GAAGGCTCATCTTTGATGTTTATCAAGGCTGATTTGTAGGT 807
RESULT 37
ACD66743
ID AC366743 standard; cDNA; 1554 BP.
XX ACD66743;
AC ACD66743;
XX 17-SEP-2003 (first entry)
XX Secreted polypeptide-related cDNA #71.
XX Mouse; Gene: ss; TANGO; INTERCEPT; secreted polypeptide; immune disorder;
XX hormonal disorder; proliferative disorder; cancer; thyroid disorder;
XX diabetes; multiple sclerosis; lupus; neurological disorder; anaemia;
XX Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
XX myocardial infarction; congestive heart disease; blood platelet disorder;
XX thrombocytopenia; blood vessel; atherosclerosis; vasculitis.
XX Mus sp.
XX US2003022279-A1.
XX 30-JAN-2003.
XX 12-JAN-2001; 2001US-00759130.
XX 14-JUN-1999; 99US-00333159.
XX 29-JUN-1999; 99US-00342364.

PR 10-SEP-1999; 99US-00393996.
PR 19-OCT-1999; 99US-00420707.
PR 07-JAN-2000; 2000US-00479249.
PR 27-APR-2000; 2000US-00559497.
PR 24-MAY-2000; 2000US-00578063.
PR 16-JUN-2000; 2000US-00596194.
PR 23-JUN-2000; 2000US-00602871.
PR 30-JUN-2000; 2000US-00608452.
XX (FRAS/) FRASER C C.
PA (BARN/) BARNES T M.
PA (SHAR/) SHARP J D.
PA (KIRS/) KIRST S J.
PA (MYER/) MYERS P S.
PA (LEIB/) LEIBY K R.
PA (HOLT/) HOLTZMAN D A.
PA (MCCA/) MCCARTHY S A.
PA (WRIGHT/) WRIGHTON N.
PA (MACK/) MACKAY C R.
PA (GOOD/) GOODEARL A D J.
XX Fraser CC, Barnes TM, Sharp JD, Kirst SJ, Myers PS, Leiby KR;
XX Holtzman DA, McCarthy SA, Wrighton N, Mackay CR, Goodearl ADJ;
XX WPI: 2003-456290/43.
XX P-PSDB; ABO32585.
XX New nucleic acid molecule encoding a secreted protein (e.g. TANGO 202,
XX TANGO 210 or INTERCEPT 217), useful for diagnosing, preventing or
XX treating disorders such as cancer, diabetes or atherosclerosis, and in
XX forensic biology.
XX PS Claim 2; SEQ ID NO 182; 482pp; English.
XX The invention relates to secreted polypeptide-related proteins and
XX nucleic acids (TANGO and INTERCEPT proteins and nucleic acids). The
XX nucleic acids, proteins and antibodies specific to the proteins are
XX useful in screening assays, predictive medicine (e.g. diagnostic assays,
XX prognostic assays, monitoring clinical trials and pharmacogenetics) and
XX prophylactic and therapeutic methods. The sequences are used in
XX diagnosing, preventing or treating proliferative disorders (e.g.
XX cancers), hormonal disorders (e.g. diabetes or thyroid disorders), immune
XX disorders (e.g. multiple sclerosis or lupus), neurological disorders
XX (e.g. Alzheimer's disease or Parkinson's disease), cardiovascular
XX disorders (e.g. myocardial infarction or congestive heart disease), blood
XX platelet disorders (e.g. thrombocytopenia or anaemia) and disorders
XX involving blood vessels (e.g. atherosclerosis or vasculitis). The nucleic
XX acids may also be used in chromosome mapping, tissue typing and forensic
XX biology, and as surrogate markers. This sequence represents a secreted
XX polypeptide-related cDNA of the invention. Note: The sequence data for
XX this patent was obtained in electronic format directly from USPTO at
XX seqdata.uspto.gov/sequence.html
XX SQ Sequence 1554 BP; 430 A; 343 C; 355 G; 426 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 210 Length: 1554
Score: 40.00 Matches: 7
Percent Similarity: 83.33% Conservative: 3
Best Local Similarity: 58.33% Mismatches: 2
Query Match: 50.63% Indels: 0
DB: 7 Gaps: 0
US-09-737-297-3 (1-16) x ACD66743 (1-1554)
QY 2 GlucySerThr***AspValTyGlnAsnIleGln 13
DB 438 CAAGGCTCACTCCACTGACGTTTACCAGGATATCCAA 473
RESULT 38
AAS97277/C
ID AAS97277 standard; DNA; 1773 BP.
XX

AC AAS97277;
CX 12-MAR-2002 (first entry)
CX Neisseria meningitidis virulence gene #82.
X Meningitis; virulence; gene; antibacterial; vaccine; veterinary;
W infection; Gram-negative bacteria; antimicrobial; ds.
CX Neisseria meningitidis.
CX WO200185772-A2.
CX 15-NOV-2001.
CX 08-MAY-2001; 2001WO-GB002003.
CX 08-MAY-2000; 2000GB-00011108.
CX (MICR-) MICROSCIENCE LTD.
CX Tang C;
CX WPI; 2002-066593/09.
CX P-PSDB; AAU72992.
CX New peptide encoded by operon including virulence genes of Neisseria
CX meningitidis, useful as vaccine component for treating or preventing
CX meningitis and for identifying antimicrobial drug.
CX Claim 1; Page 332-335; 423pp; English.
CX The invention relates to a peptide (I) encoded by an operon (II) of
CX Neisseria meningitidis including virulence genes, or a related molecule
CX having a 40% sequence similarity at the peptide or nucleotide level in a
CX Gram-negative bacterium, or its functional fragment, for therapeutic or
CX diagnostic use. (I) and (II) are useful in the manufacture of a
CX medicament for treating or preventing a condition (e.g., meningitis)
CX associated with infection by Neisseria or Gram-negative bacteria. The
CX product is useful for veterinary treatment and in a screening assay for
CX the identification of an antimicrobial drug. The vaccines have
CX prophylactic applications. AAS97196-AAS97305 represent N. meningitidis
CX virulence genes and related PCR primers of the invention
CX Sequence 1773 BP; 418 A; 528 C; 498 G; 329 T; 0 U; 0 Other;
SQ Alignment Scores: 247 Length: 1773
Pred. No.: 40.00 Matches: 8
Score: 62.50% Conservative: 2
Percent Similarity: 50.00% Mismatches: 6
Best Local Similarity: 50.63% Indels: 0
Query Match: 6 Gaps: 0
DB: 6
US-09-737-297-3 (1-16) x AAS97277 (1-1773)
Qy 1 AlaGluGlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16
Db 1506 GCGGTTGGCGGTACGGATGAGGTGTTGCAGCATTTTCAGCAGCGGG 1459
RESULT 39
AAS97234/C
ID AAS97234 standard; DNA; 1773 BP.
AC AAS97234;
CX 12-MAR-2002 (first entry)
CX Neisseria meningitidis virulence gene #39.
CX Meningitis; virulence; gene; antibacterial; vaccine; veterinary;
KW infection; Gram-negative bacteria; antimicrobial; ds.
CX

OS Neisseria meningitidis.
CX WO200185772-A2.
CX 15-NOV-2001.
CX 08-MAY-2001; 2001WO-GB002003.
CX 08-MAY-2000; 2000GB-00011108.
CX (MICR-) MICROSCIENCE LTD.
CX Tang C;
CX WPI; 2002-066593/09.
CX P-PSDB; AAU72949.
CX New peptide encoded by operon including virulence genes of Neisseria
CX meningitidis, useful as vaccine component for treating or preventing
CX meningitis and for identifying antimicrobial drug.
CX Claim 1; Page 164-167; 423pp; English.
CX The invention relates to a peptide (I) encoded by an operon (II) of
CX Neisseria meningitidis including virulence genes, or a related molecule
CX having a 40% sequence similarity at the peptide or nucleotide level in a
CX Gram-negative bacterium, or its functional fragment, for therapeutic or
CX diagnostic use. (I) and (II) are useful in the manufacture of a
CX medicament for treating or preventing a condition (e.g., meningitis)
CX associated with infection by Neisseria or Gram-negative bacteria. The
CX product is useful for veterinary treatment and in a screening assay for
CX the identification of an antimicrobial drug. The vaccines have
CX prophylactic applications. AAS97196-AAS97305 represent N. meningitidis
CX virulence genes and related PCR primers of the invention
CX Sequence 1773 BP; 418 A; 528 C; 498 G; 329 T; 0 U; 0 Other;
SQ Alignment Scores: 247 Length: 1773
Pred. No.: 40.00 Matches: 8
Score: 62.50% Conservative: 2
Percent Similarity: 50.00% Mismatches: 6
Best Local Similarity: 50.63% Indels: 0
Query Match: 6 Gaps: 0
DB: 6
US-09-737-297-3 (1-16) x AAS97234 (1-1773)
Qy 1 AlaGluGlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16
Db 1506 GCGGTTGGCGGTACGGATGAGGTGTTGCAGCATTTTCAGCAGCGGG 1459
RESULT 40
AAS97435/C
ID AAS97435 standard; DNA; 1773 BP.
CX AAS97435;
CX 27-FEB-2002 (first entry)
CX N. meningitidis phli gene DNA sequence.
CX DNA uptake; veterinary; antiinflammatory; antibiotic; antibacterial;
CX vaccine; phli; ds.
CX Neisseria meningitidis.
CX Key Location/Qualifiers
CX CDS 1..1773
CX FT /*tag= a
CX FT /gene= "phli"
CX WO200187939-A2.
CX

PD 22-NOV-2001.
XX
PF 18-MAY-2001; 2001WO-GB002247.
XX
XX 08-FEB-2002; 2002US-00072851.
XX
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX

PI Tang C;
XX
XX WPI; 2002-041720/05.
DR P-PSDB; AAG66072.
DR
XX

XX New polypeptide useful as vaccine for immunizing animals against
PT bacterial infections, is encoded by genes from *Neisseria meningitidis* and
PT polynucleotides for obtaining microorganisms having reduced ability to
PT uptake DNA.
XX

PS Claim 10; Page 24-27; 55pp; English.
XX

XX The invention provides a series of genes from *N. meningitidis* that are
CC shown to encode products responsible for DNA uptake. A transformation
CC deficient microorganism comprising a mutation that disrupts the
CC expression of a gene, a vaccine comprising such a microorganism, the
CC genes and their products are useful for manufacture of medicament for use
CC in treatment or prevention of a condition associated with infection by *N.*
CC *meningitidis* or Gram-negative bacteria e.g. *meningitis* for veterinary
CC treatment. The polynucleotide sequences are useful for searching related
CC genes or peptides in other microorganisms. The peptides encoded by the
CC genes are useful for the preparation of antibodies which is used in
CC passive immunization or in diagnostic applications. The transformation
CC deficient microorganism is useful as a carrier system for the delivery of
CC heterologous antigens, therapeutic proteins or nucleic acids in vivo. The
CC present sequence represents a *N. meningitidis* phli gene DNA having a
CC phosphoenolpyruvate-protein phosphotransferase function
XX

SQ Sequence 1773 BP; 418 A; 528 C; 498 G; 329 T; 0 U; 0 Other;
XX

Alignment Scores:
Pred. No.: 247 Length: 1773
Score: 40.00 Matches: 8
Percent Similarity: 62.50% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 50.63% Indels: 0
DB: 6 Gaps: 0

US-09-737-297-3 (1-16) x AAF167435 (1-1773)

OY 1 AlagluGlySerThr***AspValTyrglnAsnIleGlnTyAlaGly 16
DB 1506 GCGTTGGCGGTACGATGACGTGTTCAGCATTTTCACACACGCGGG 1459

RESULT 41
ACA41645/c
ID ACA41645 standard; DNA; 1776 BP.
XX

AC ACA41645;
XX

DT 19-JUN-2003 (first entry)
XX

DE Prokaryotic essential gene #23302.
XX

XX Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX

OS *Neisseria meningitidis*.
XX

PN W020027183-A2.
XX

PD 03-OCT-2002.
XX

PF 21-MAR-2002; 2002WO-US0009107.
XX

PR 21-MAR-2001; 2001US-00915242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
XX
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu RH;
XX
XX WPI; 2003-029926/02.
DR P-PSDB; ABU37775.
DR

XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX

PS Claim 14; SEQ ID NO 29515; 1766pp; English.
XX

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway;
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 1776 BP; 420 A; 528 C; 498 G; 330 T; 0 U; 0 Other;
XX

Alignment Scores:
Pred. No.: 247 Length: 1776
Score: 40.00 Matches: 8
Percent Similarity: 62.50% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 50.63% Indels: 0
DB: 7 Gaps: 0

US-09-737-297-3 (1-16) x ACA41645 (1-1776)

OY 1 AlagluGlySerThr***AspValTyrglnAsnIleGlnTyAlaGly 16
DB 1506 GCGTTGGCGGTACGATGACGTGTTCAGCATTTTCACACACGCGGG 1459

RESULT 42

AAF16209/c

ID AAF16209 standard; cDNA; 1793 BP.

XX

AC AAF16209;

X T 13-MAR-2001 (first entry)
X Human prostate cancer antigen nucleotide sequence SEQ ID NO:644.
X Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
W neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
W vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
W antibacterial; gene therapy; neural; immune; reproductive; renal;
W gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
W wound; infectious disease; ss.
X Homo sapiens.
X WO200055174-A1.
X 21-SEP-2000.
X 08-MAR-2000; 2000WO-US005988.
X 12-MAR-1999; 99US-0124270P.
X (HUMA-) HUMAN GENOME SCI INC.
X (ROSE/) ROSEN C A.
X Rosen CA, Ruben SM;
X WPI: 2000-587513/55.
X P-PSDB; AAB57006.
X Prostate cancer associated gene sequences, referred to as prostate cancer
T antigens, useful for treatment, prevention, and diagnosis of disorders
T such as prostate cancer.
X Claim 1; Page 1086; 2338pp; English.
X AAF15566 to AAF16505 encode the human prostate cancer associated
C proteins, called prostate cancer antigens, given in AAB56333 to AAB57302.
C The prostate cancer antigens can have neuroprotective, cytostatic,
C cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
C nephrotropic, antiinfective, gynaecological and antibacterial activities,
C and can be used in gene therapy. The prostate cancer antigen
C polynucleotides may be used for detection of prostate cancer, chromosome
C identification, as chromosome markers, and for numerous other diagnostic
C or research purposes. The prostate cancer antigens may be used to treat
C disorders such as neural, immune, muscular, reproductive,
C gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
C disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
C AAB57303 represent sequences used in the exemplification of the present
C invention
X X Sequence 1793 BP; 570 A; 315 C; 396 G; 506 T; 0 U; 6 Other;
X Alignment Scores:
Pred. No.: 250 Length: 1793
Score: 40.00 Matches: 6
Percent Similarity: 90.91% Conservative: 4
Best Local Similarity: 54.55% Mismatches: 1
Query Match: 50.63% Indels: 0
DB: 3 Gaps: 0
US-09-737-297-3 (1-16) x AAF16209 (1-1793)
QY 5 Thr***AspValTyGlnAsnIleGlnTyrAla 15
Db 1668 ACTGAGATATTATAGACTCACTCATATGCT 1636
RESULT 43
AAH22890/c
ID AAH22890 standard; DNA; 1812 BP.
XX AAH22890;
XX

DT 17-SEP-2001 (first entry)
XX R. ruber cyclododecanone monooxygenase cdda gene homologue.
DS Cyclododecanone; degradation; lauryl lactone esterase; cdda; cddb; cddx;
KW dodecanoic diacid; cyclododecanone monooxygenase; cyclic ketone; cdd; cddx;
KW 12-hydroxylauric acid dehydrogenase; 12-oxo lauric acid dehydrogenase;
KW macrolactone; microbiological; cddc; cdd; ds.
OS Rhodococcus ruber.
XX WO200142436-A2.
XX 14-JUN-2001.
XX 08-DEC-2000; 2000WO-US033426.
XX 10-DEC-1999; 99US-0170214P.
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX Chen MW, Cheng Q, Gibson KJ, Kostichka KN, Thomas SM;
XX WPI: 2001-451630/48.
XX Isolated nucleic acids encoding dodecanoic diacid synthesizing enzyme,
PT cyclododecanone monooxygenase for bioproduction of dodecanoic diacid from
PT cyclododecanone.
XX Claim 8; Page 70; 78pp; English.
XX The invention relates to genes from Rhodococcus ruber involved in
CC cyclododecanone degradation pathway. A 10 kb cluster from R. ruber was
CC isolated that comprises genes (designated cdda, cddb, cddx, cddy, cddc
CC and cdd) encoding enzymes such as dodecanoic diacid synthesizing enzyme,
CC cyclododecanone monooxygenase, lauryl lactone esterase, 12-hydroxylauric
CC acid dehydrogenase, 12-oxo lauric acid dehydrogenase. Host cells
CC comprising the nucleic acids are used for the production of dodecanoic
CC acid when contacted with cyclododecanone, lauryl lactone when contacted
CC with cyclododecanone, 12-hydroxy lauric acid when contacted with lauryl
CC lactone, 12-oxo lauric acid when contacted with 12-hydroxy lauric acid
CC and dodecanoic diacid when contacted with 12-oxo lauric acid. They are
CC also useful for the production of hydroxy acids when contacted with
CC cyclic ketones with 6 to eight carbon atoms like C6, C10, C11, C12, C13
CC and C15 cyclic ketones, cyclohexanone, cyclodecanone, cycloundecanone,
CC cyclo dodecanone, cyclotridecanone and cyclopentadecanone. The host cells
CC are also used for the production of macrolactones with at least 10 carbon
CC atoms when contacted with a cyclic ketone. The nucleic acids are useful
CC for the bioproduction of dodecanoic diacid from cyclododecanone by
CC microbiological means. Sequences AAB22888-90 represent nucleic acid
CC sequences which will hybridise under stringent conditions to the cdda
CC gene from R. ruber encoding a cyclododecanone monooxygenase enzyme
XX X Sequence 1812 BP; 384 A; 591 C; 551 G; 286 T; 0 U; 0 Other;
X Alignment Scores:
Pred. No.: 253 Length: 1812
Score: 40.00 Matches: 7
Percent Similarity: 68.75% Conservative: 4
Best Local Similarity: 43.75% Mismatches: 5
Query Match: 50.63% Indels: 0
DB: 4 Gaps: 0
US-09-737-297-3 (1-16) x AAH22890 (1-1812)
QY 1 AlaGluGlySerThr***AspValTyGlnAsnIleGlnTyrAlaGly 16
Db 806 GCCGAGCGGTGAGTTGGACTACGTACCGATCCGCTCAGTACGTGGC 759
RESULT 44
AAH18171/c
ID AAH18171 standard; CDNA; 1927 BP.
XX

ACD66742 standard; cDNA; 2467 BP.
ACD66742;
17-SEP-2003 (first entry)
Secreted polypeptide-related cDNA #70.
Mouse; gene; ss; TANGO; INTERCEPT; secreted polypeptide; immune disorder;
hormonal disorder; proliferative disorder; cancer; thyroid disorder;
diabetes; multiple sclerosis; lupus; neurological disorder; anaemia;
Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
myocardial infarction; congestive heart disease; blood platelet disorder;
thrombocytopenia; blood vessel; atherosclerosis; vasculitis.
Mus sp.
US2003022279-A1.
30-JAN-2003.
12-JAN-2001; 2001US-00759130.
14-JUN-1999; 99US-00333159.
29-JUN-1999; 99US-00342364.
10-SEP-1999; 99US-00393996.
19-OCT-1999; 99US-00420707.
07-JAN-2000; 2000US-00479249.
27-APR-2000; 2000US-00559497.
24-MAY-2000; 2000US-00578063.
16-JUN-2000; 2000US-00596194.
23-JUN-2000; 2000US-00602871.
30-JUN-2000; 2000US-00609452.
(FRASER) FRASER C C.
(BARN) BARNES T M.
(SHAR) SHARP J D.
(KIRS) KIRST S J.
(MYER) MYERS P S.
(LEIB) LEIBY K R.
(HOLT) HOLTZMAN D A.
(MCCA) MCCARTHY S A.
(WRIG) WRIGHTON N.
(MACK) MACKAY C R.
(GOOD) GOODEARL A D J.
Fraser CC, Barnes TM, Sharp JD, Kirst SJ, Myers PS, Leiby KR;
Holtzman DA, McCarthy SA, Wrighton N, Mackay CR, Goodearl ADJ;
P-PSDB; ABO32585.
New nucleic acid molecule encoding a secreted protein (e.g. TANGO 202,
TANGO 210 or INTERCEPT 217), useful for diagnosing, preventing or
treating disorders such as cancer, diabetes or atherosclerosis, and in
forensic biology.
Claim 2; Fig 15F-15I; 482pp; English.
The invention relates to secreted polypeptide-related proteins and
nucleic acids (TANGO and INTERCEPT proteins and nucleic acids). The
nucleic acids, proteins and antibodies specific to the proteins are
useful in screening assays, predictive medicine (e.g. diagnostic assays,
prognostic assays, monitoring clinical trials and pharmacogenetics) and
prophylactic and therapeutic methods. The sequences are used in
diagnosing, preventing or treating proliferative disorders (e.g.
cancers), hormonal disorders (e.g. diabetes or thyroid disorders), immune
disorders (e.g. multiple sclerosis or lupus), neurological disorders
(e.g. Alzheimer's disease or Parkinson's disease), cardiovascular
disorders (e.g. myocardial infarction or congestive heart disease), blood
platelet disorders (e.g. thrombocytopenia or anaemia) and disorders
involving blood vessels (e.g. atherosclerosis or vasculitis). The nucleic
acids may also be used in chromosome mapping, tissue typing and forensic

CC biology, and as surrogate markers. This sequence represents a secreted
CC polypeptide-related cDNA of the invention. Note: The sequence data for
CC this patent was obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 2467 BP; 792 A; 533 C; 493 G; 649 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 370 Length: 2467
Score: 40.00 Matches: 7
Percent Similarity: 83.33% Conservative: 3
Best Local Similarity: 58.33% Mismatches: 2
Query Match: 50.63% Indels: 0
DB: 7 Gaps: 0
US-09-737-297-3 (1-16) x ACD66742 (1-2467)
QY 2 GluglySerThr***AspValTyrGlnAsnIleGln 13
Db 438 CMAAGTCACTCCACTGACGTTTACCGATATCCAA 473
RESULT 50
ABL29911/c
ID ABL29911 standard; DNA; 3303 BP.
XX
AC ABL29911;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 41206.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX
OS Drosophila melanogaster.
XX
DN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
XX
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 41206; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737,
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 3303 BP; 904 A; 926 C; 853 G; 620 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 530 Length: 3303
Score: 40.00 Matches: 7

Percent Similarity: 69.23%
 Best Local Similarity: 53.85%
 Query Match: 50.63%
 DB: 4

Conservative: 2
 Mismatches: 4
 Indels: 0
 Gaps: 0

JS-09-737-297-3 (1-16) x ABL29911 (1-3303)

2y 2 GluGlySerThr***AspValTyrGlnAsnIleGlnTyr 14
 54 GAGGTACCAAGATGAATGTATATACAAATTTCCATAT 16

Search completed: March 1, 2004, 10:23:45
 Job time : 397 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - nucleic search, using frame_plus_p2n model

un on: March 1, 2004, 10:15:45 ; Search time 2489 Seconds
(without alignments)
191.963 Million cell updates/sec

title: US-09-737-297-3

effect score: 79

sequence: 1 AEGSTXDVTQNIQVAG 16

coring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

searched: 27513289 seqs, 14931090276 residues

total number of hits satisfying chosen parameters: 55026578

minimum DB seq length: 0

maximum DB seq length: 2000000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlp

-Q=/cgn2_1/USPTO_spool_p/US09737297/runat_01032004_085323_3469/app.query.fasta_1.199

-DB=EST -QPMF=fastac -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0

-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=150

-DOALIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=50 -MODE=LOCAL

-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USBR=US09737297@cgn 1.1 4237 @runat_01032004_085323_3469 -ICPU=3

-NO MMAP -LARGEQUERY -NEG SCORES -WAIT -DSPBLOCK=100 -LONGLOG

-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:

1: em_estba:

2: em_esthum:

3: em_estin:

4: em_estnu:

5: em_estov:

6: em_estpl:

7: em_estro:

8: em_hcti:

9: gb_est1:

10: gb_est2:

11: gb_hcti:

12: gb_est3:

13: gb_est4:

14: gb_est5:

15: em_estfun:

16: em_estom:

17: em_gss_hum:

18: em_gss_inv:

19: em_gss_pln:

20: em_gss_vrt:

21: em_gss_fut:

22: em_gss_mam:

23: em_gss_mus:

24: em_gss_pro:

25: em_gss_rod:

26: em_gss_pbg:

27: em_gss_vri:

28: gb_gss1:

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	50	63.3	1003	10	BE789874	BE789874 601476741
2	48	60.8	464	12	BJ341043	BJ341043 BJ341043
3	48	60.8	570	12	BI979979	BI979979 ft78d09.x
4	48	60.8	675	12	BJ402251	BJ402251 BJ402251
5	48	60.8	681	12	BJ373041	BJ373041 BJ373041
6	48	60.8	681	12	BJ375875	BJ375875 BJ375875
7	48	60.8	681	12	BJ402833	BJ402833 BJ402833
8	48	60.8	683	12	BJ399741	BJ399741 BJ399741
9	48	60.8	692	12	BJ400641	BJ400641 BJ400641
10	48	60.8	702	12	BJ377938	BJ377938 BJ377938
11	48	60.8	704	12	BJ374091	BJ374091 BJ374091
12	48	60.8	728	12	BJ371887	BJ371887 BJ371887
13	48	60.8	733	12	BJ400719	BJ400719 BJ400719
14	46	58.2	414	28	AQ014772	AQ014772 HS_2176.A
15	45	57.0	567	14	CD598562	CD598562 EST15085
16	45	57.0	775	14	CKJ41671	CKJ41671 AGENCOURT
17	44	55.7	171	9	AU054137	AU054137 AU054137
18	44	55.7	279	9	AU034493	AU034493 AU034493
19	44	55.7	306	29	CE760053	CE760053 tigr-gss-
20	44	55.7	326	9	AU038380	AU038380 AU038380
21	44	55.7	360	9	AV191946	AV191946 AV191946
22	44	55.7	360	9	AV194943	AV194943 AV194943
23	44	55.7	360	9	AV195613	AV195613 AV195613
24	44	55.7	370	9	AV669128	AV669128 AV669128
25	44	55.7	373	9	AU034232	AU034232 AU034232
26	44	55.7	376	13	C71480	C71480 C71480 Yxji
27	44	55.7	388	10	BF047887	BF047887 dc89g03.y
28	44	55.7	415	12	BJ034056	BJ034056 BJ034056
29	44	55.7	418	29	CC469417	CC469417 CH240.142
30	44	55.7	438	10	BG037933	BG037933 dc59h11.y
31	44	55.7	456	13	CB3991	CB3991 CB3991 Dict
32	44	55.7	457	12	BG408335	BG408335 dc98g09.y
33	44	55.7	477	12	BJ070599	BJ070599 BJ070599
34	44	55.7	481	10	BF426652	BF426652 df71d05.y
35	44	55.7	482	10	BE506936	BE506936 db66f10.y
36	44	55.7	489	9	AU054046	AU054046 AU054046
37	44	55.7	518	9	AU034593	AU034593 AU034593
38	44	55.7	519	28	BZ991276	BZ991276 PUDDS42TD
39	44	55.7	533	28	AQ405209	AQ405209 HS_5038.B
40	44	55.7	535	13	C91372	C91372 C91372 Dict
41	44	55.7	539	12	BM879901	BM879901 ku02d02.y
42	44	55.7	541	12	BG226255	BG226255 kgj9d02.y
43	44	55.7	549	9	AU053352	AU053352 AU053352
44	44	55.7	559	9	AU053355	AU053355 AU053355
45	44	55.7	562	9	AJ272705	AJ272705 AJ272705
46	44	55.7	563	12	BG486999	BG486999 dc97c11.y
47	44	55.7	569	9	AU060911	AU060911 AU060911
48	44	55.7	569	9	AA522323	AA522323 vi45c11.x
49	44	55.7	569	13	C94080	C94080 C94080 Dict
50	44	55.7	607	13	C93278	C93278 C93278 Dict
51	44	55.7	623	28	AZ159123	AZ159123 SP_0062.A
52	44	55.7	625	13	C92830	C92830 C92830 Dict
53	44	55.7	630	14	CB199895	CB199895 AGENCOURT
54	44	55.7	633	9	AU0200196	AU0200196 AU0200196
55	44	55.7	641	12	BJ034364	BJ034364 BJ034364
56	44	55.7	642	9	AU033714	AU033714 AU033714
57	44	55.7	653	13	C92791	C92791 C92791 Dict
58	44	55.7	662	9	AU051981	AU051981 AU051981
59	44	55.7	672	9	AU038147	AU038147 AU038147
60	44	55.7	676	12	BJ401761	BJ401761 BJ401761
61	44	55.7	681	13	C92664	C92664 C92664 Dict
62	44	55.7	688	9	AU039255	AU039255 AU039255

40 ACCTGGATGTTTATCAAAATATTCAATAC 69

RESULT 6
 BJ375875 681 bp mRNA linear EST 08-MAR-2002
 LOCUS Dictyostelium discoideum cDNA library, CF Dictyostelium
 DEFINITION dictyostelium cDNA clone ddc20107 3', mRNA sequence.

ACCESSION BJ375875
 VERSION BJ375875.1 GI:19285258
 KEYWORDS EST
 SOURCE Dictyostelium discoideum
 ORGANISM Dictyostelium discoideum
 Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 REFERENCE 1 (bases 1 to 681)
 AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
 TITLE Full length cDNA of Dictyostelium discoideum at the culmination stage

JOURNAL Unpublished (2002)
 COMMENT Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp.

FEATURES
 source Location/Qualifiers
 1..681
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="ddc20107"
 /sex="mat A"
 /dev_stage="Culmination stage"
 /clone_lib="Dictyostelium discoideum cDNA library, CF"

ORIGIN
 Alignment Scores:
 Pred. No.: 90.7 Length: 681
 Score: 48.00 Matches: 9
 Percent Similarity: 90.00% Conservative: 0
 Best Local Similarity: 90.00% Mismatches: 1
 Query Match: 90.76% Indels: 0
 DB: 12 Gaps: 0

US-09-737-297-3 (1-16) x BJ375875 (1-681)

Qy 5 Thr***AspValtyrGlnAsnIleGlnTyr 14
 |||||
 40 ACCTGGATGTTTATCAAAATATTCAATAC 69

DB

RESULT 7
 BJ402833 681 bp mRNA linear EST 10-MAR-2002
 LOCUS Dictyostelium discoideum cDNA library, SF Dictyostelium
 DEFINITION dictyostelium cDNA clone dds18m07 3', mRNA sequence.

ACCESSION BJ402833
 VERSION BJ402833.1 GI:19315750
 KEYWORDS EST
 SOURCE Dictyostelium discoideum
 ORGANISM Dictyostelium discoideum
 Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 REFERENCE 1 (bases 1 to 681)
 AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
 TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp.

```
FEATURES
source
  Location/Qualifiers
    1..681
    /organism="Dictyostelium discoideum"
    /mol_type="mrna"
    /strain="AX4"
    /db_xref="taxon:44689"
    /clone="ddsl8m07"
    /sex="mat A"
    /dev_stage="slug stage"
    /clone_lib="Dictyostelium discoideum cDNA library, SF"
ORIGIN
Alignment Scores:
  Pred. No.: 90.7 Length: 681
  Score: 48.00 Matches: 9
  Percent Similarity: 90.00% Conservative: 0
  Best Local Similarity: 90.00% Mismatches: 1
  Query Match: 60.76% Indels: 0
  DB: 12 Gaps: 0
US-09-737-297-3 (1-16) x BJ402833 (1-681)
Y 5 Thr***AspValtyrGlnAsnIleGlnTyr 14
  |||
  41 ACCTTGGATGTTTATCAAAATATTCAATAC 70
RESULT 8
BJ399741
LOCUS BJ399741 Dictyostelium discoideum cDNA library, SF Dictyostelium
DEFINITION dictyostelium cDNA clone dds7409 3', mRNA sequence.
ACCESSION BJ399741
VERSION BJ399741.1 GI:19312658
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
  1..683
  /organism="Dictyostelium discoideum"
  /mol_type="mrna"
  /strain="AX4"
  /db_xref="taxon:44689"
  /clone="ddsl8m07"
  /sex="mat A"
  /dev_stage="slug stage"
  /clone_lib="Dictyostelium discoideum cDNA library, SF"
ORIGIN
Alignment Scores:
  Pred. No.: 91.1 Length: 683
  Score: 48.00 Matches: 9
  Percent Similarity: 90.00% Conservative: 0
  Best Local Similarity: 90.00% Mismatches: 1
  Query Match: 60.76% Indels: 0
  DB: 12 Gaps: 0
US-09-737-297-3 (1-16) x BJ399741 (1-683)
Y 5 Thr***AspValtyrGlnAsnIleGlnTyr 14
  |||
  44 ACCTTGGATGTTTATCAAAATATTCAATAC 73
FEATURES
source
  Location/Qualifiers
    1..683
    /organism="Dictyostelium discoideum"
    /mol_type="mrna"
    /strain="AX4"
    /db_xref="taxon:44689"
    /clone="ddsl8m07"
    /sex="mat A"
    /dev_stage="slug stage"
    /clone_lib="Dictyostelium discoideum cDNA library, SF"
ORIGIN
Alignment Scores:
  Pred. No.: 91.1 Length: 683
  Score: 48.00 Matches: 9
  Percent Similarity: 90.00% Conservative: 0
  Best Local Similarity: 90.00% Mismatches: 1
  Query Match: 60.76% Indels: 0
  DB: 12 Gaps: 0
US-09-737-297-3 (1-16) x BJ399741 (1-683)
Y 5 Thr***AspValtyrGlnAsnIleGlnTyr 14
  |||
  44 ACCTTGGATGTTTATCAAAATATTCAATAC 73
```

```
RESULT 9
BJ400641
LOCUS BJ400641 Dictyostelium discoideum cDNA library, SF Dictyostelium
DEFINITION dictyostelium cDNA clone dds14k01 3', mRNA sequence.
ACCESSION BJ400641
VERSION BJ400641.1 GI:19313558
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
  1..692
  /organism="Dictyostelium discoideum"
  /mol_type="mrna"
  /strain="AX4"
  /db_xref="taxon:44689"
  /clone="ddsl4k01"
  /sex="mat A"
  /dev_stage="slug stage"
  /clone_lib="Dictyostelium discoideum cDNA library, SF"
ORIGIN
Alignment Scores:
  Pred. No.: 92.7 Length: 692
  Score: 48.00 Matches: 9
  Percent Similarity: 90.00% Conservative: 0
  Best Local Similarity: 90.00% Mismatches: 1
  Query Match: 60.76% Indels: 0
  DB: 12 Gaps: 0
US-09-737-297-3 (1-16) x BJ400641 (1-692)
Y 5 Thr***AspValtyrGlnAsnIleGlnTyr 14
  |||
  47 ACCTTGGATGTTTATCAAAATATTCAATAC 76
FEATURES
source
  Location/Qualifiers
    1..692
    /organism="Dictyostelium discoideum"
    /mol_type="mrna"
    /strain="AX4"
    /db_xref="taxon:44689"
    /clone="ddsl4k01"
    /sex="mat A"
    /dev_stage="slug stage"
    /clone_lib="Dictyostelium discoideum cDNA library, SF"
ORIGIN
Alignment Scores:
  Pred. No.: 92.7 Length: 692
  Score: 48.00 Matches: 9
  Percent Similarity: 90.00% Conservative: 0
  Best Local Similarity: 90.00% Mismatches: 1
  Query Match: 60.76% Indels: 0
  DB: 12 Gaps: 0
US-09-737-297-3 (1-16) x BJ400641 (1-692)
Y 5 Thr***AspValtyrGlnAsnIleGlnTyr 14
  |||
  47 ACCTTGGATGTTTATCAAAATATTCAATAC 76
FEATURES
source
  Location/Qualifiers
    1..692
    /organism="Dictyostelium discoideum"
    /mol_type="mrna"
    /strain="AX4"
    /db_xref="taxon:44689"
    /clone="ddsl4k01"
    /sex="mat A"
    /dev_stage="slug stage"
    /clone_lib="Dictyostelium discoideum cDNA library, SF"
ORIGIN
Alignment Scores:
  Pred. No.: 92.7 Length: 692
  Score: 48.00 Matches: 9
  Percent Similarity: 90.00% Conservative: 0
  Best Local Similarity: 90.00% Mismatches: 1
  Query Match: 60.76% Indels: 0
  DB: 12 Gaps: 0
US-09-737-297-3 (1-16) x BJ400641 (1-692)
Y 5 Thr***AspValtyrGlnAsnIleGlnTyr 14
  |||
  47 ACCTTGGATGTTTATCAAAATATTCAATAC 76
```

source
1. .702
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc26b13"
/sex="mat A"
/dev_stage="Culmination stage"
/clone_lib="Dictyostelium discoideum cDNA library, CF"
ORIGIN
Alignment Scores:
Pred. No.: 94.5 Length: 702
Score: 48.00 Matches: 9
Percent Similarity: 90.00% Conservative: 0
Best Local Similarity: 90.00% Mismatches: 1
Query Match: 60.76% Indels: 0
Gaps: 0
US-09-737-297-3 (1-16) x BJ377938 (1-702)
QY 5 Thr***AspValtyrGlnAsnIleGlnTyr 14
DB 41 ACCTGGATGTTTATCAAAATATTCATAC 70

RESULT 11
BJ374091
LOCUS BJ374091 704 bp mRNA linear EST 08-MAR-2002
DEFINITION BJ374091 Dictyostelium discoideum cDNA library, CF Dictyostelium
discoideum cDNA clone ddc6p10 3', mRNA sequence.
ACCESSION BJ374091
VERSION BJ374091.1 GI:19283474
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 704)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the culmination
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
source
1. .704
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc6p10"
/sex="mat A"
/dev_stage="Culmination stage"
/clone_lib="Dictyostelium discoideum cDNA library, CF"

ORIGIN
Alignment Scores:
Pred. No.: 94.9 Length: 704
Score: 48.00 Matches: 9
Percent Similarity: 90.00% Conservative: 0
Best Local Similarity: 90.00% Mismatches: 1
Query Match: 60.76% Indels: 0
Gaps: 0
US-09-737-297-3 (1-16) x BJ374091 (1-704)
QY 5 Thr***AspValtyrGlnAsnIleGlnTyr 14
DB 41 ACCTGGATGTTTATCAAAATATTCATAC 70

RESULT 12
BJ371887
LOCUS BJ371887 728 bp mRNA linear EST 08-MAR-2002
DEFINITION BJ371887 Dictyostelium discoideum cDNA library, CF Dictyostelium
discoideum cDNA clone ddc10m03 3', mRNA sequence.
ACCESSION BJ371887
VERSION BJ371887.1 GI:19281270
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 728)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the culmination
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
source
1. .728
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc10m03"
/sex="mat A"
/dev_stage="Culmination stage"
/clone_lib="Dictyostelium discoideum cDNA library, CF"

ORIGIN
Alignment Scores:
Pred. No.: 99.3 Length: 728
Score: 48.00 Matches: 9
Percent Similarity: 90.00% Conservative: 0
Best Local Similarity: 90.00% Mismatches: 1
Query Match: 60.76% Indels: 0
Gaps: 0
US-09-737-297-3 (1-16) x BJ371887 (1-728)
QY 5 Thr***AspValtyrGlnAsnIleGlnTyr 14
DB 41 ACNTTGGATGTTTATCAAAATATTCATAC 70

RESULT 13
BJ400719
LOCUS BJ400719 733 bp mRNA linear EST 10-MAR-2002
DEFINITION BJ400719 Dictyostelium discoideum cDNA library, SF Dictyostelium
discoideum cDNA clone dds14p10 3', mRNA sequence.
ACCESSION BJ400719
VERSION BJ400719.1 GI:19313636
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 733)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
Location/Qualifiers

Percent Similarity: 73.33% Conservative: 3
Best Local Similarity: 53.33% Mismatches: 4
Query Match: 58.23% Indels: 0
DB: DB: Gaps: 0

US-09-737-297-3 (1-16) x AQ014772 (1-414)

QY 1 AlaGluGlySerThr***AspValTyrGlnAsnIleGIntyAla 15
:::|::|::|::|::|::|::|::|::|::|
DDB 322 AGTGAGGTGAAGCTCTAACAATCTACACTAACTCTCAATATGCC 366

RESULT 15
CD698562 EST1508562 567 bp mRNA linear EST 25-JUN-2003
LOCUS human nasopharynx Homo sapiens cDNA, mRNA sequence.
DEFINITION CD698562
ACCESSION CD698562
VERSION 1 GI:32227011
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 567)
AUTHORS Liu,X.-Q., Zhou,Y., Zhang,L.-J., Xu,H., Chen,H.-K., Pan,Z.-G. and Zeng Y.-X.

TITLE Transcriptional Gene Expression Profile of Human Nasopharynx
JOURNAL Unpublished (2003)
COMMENT Contact: Yixin Zeng
Cancer Center
Sun Yat-sen University
651 Dongfeng Road East, Guangzhou 510060, China
Tel: 86-1380-9770-743
Fax: 86-20-8775-4506
Email: yxzeng@zsums.edu.cn.

FEATURES
source Location/Qualifiers
1..567
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue type="normal nasopharynx"
/clone_lib="human nasopharynx"
/notes="ESTs generated from a normal nasopharynx cDNA library from southern Chinese"

ORIGIN
Alignment Scores:
257 Length: 567
Pred. No.: Matches: 9
Score: 45.00
Percent Similarity: 64.29% Conservative: 0
Best Local Similarity: 64.29% Mismatches: 5
Query Match: 56.96% Indels: 0
DB: DB: Gaps: 0

US-09-737-297-3 (1-16) x CD698562 (1-567)

QY 1 AlaGluGlySerThr***AspValTyrGlnAsnIleGIntyR 14
:::|::|::|::|::|::|::|::|::|::|
DDB 22 GCCGGGGTCACAAATGACAGATGCTACATGAATATTTCCATAT 63

RESULT 16
CK141671/c EST1508562 775 bp mRNA linear EST 03-DEC-2003
LOCUS human nasopharynx Danio rerio cDNA clone IMAGE:7048054
DEFINITION AGENCOURT_16821982 NIH_ZGC_10 Danio rerio
5', mRNA sequence.
ACCESSION CK141671
VERSION CK141671.1 GI:38646968
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 775)

AUTHORS TITLE JOURNAL COMMENT

NIH-MGC <http://mgc.mci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabs-remail.nih.gov
Tissue Procurement: Len Zon, Harvard
cDNA Library Preparation: Open Biosystems
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM14815 row: b column: 20
High quality sequence stop: 750.
Location/Qualifiers

FEATURES source

1. .775
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:7048054"
/tissue_type="whole body"
/lab_host="DH10B"
/clone_lib="NH_ZGC_10"
/note="Vector: pExpress; Site 1: NotI; Site 2: EcoRV;
Bulk tissue was collected from a whole adult individual
from the Tuebingen strain. 1st strand cDNA was primed with
a Not I - oligo(dT) primer, double-stranded cDNA was
cloned into the Not I and EcoRV sites of pExpress-1.
Library was size-selected for >1 kb fragments. A
normalized version of this library is also available
(NIH_ZGC 7). Library was constructed by Open Biosystems
(Huntsville, AL)."

ORIGIN

Alignment Scores:
Pred. No.: 392 Length: 775
Score: 45.00 Matches: 7
Percent Similarity: 90.00% Conservative: 2
Best Local Similarity: 70.00% Mismatches: 1
Query Match: 56.96% Indels: 0
DB: 14 Gaps: 0

US-09-737-297-3 (1-16) x CK141671 (1-775)

Qy 7 AspValtyrGlnAsnIleGlnTy 16
Db 331 GACATCTATAGGACATACATCTGTGGA 302

RESULT 17 AU054137/c

LOCUS AU054137 Dictyostelium discoidium SL (H.Urushihara) Dictyostelium
DEFINITION AU054137 Dictyostelium discoidium SL (H.Urushihara) Dictyostelium
ACCESSION AU054137
VERSION AU054137.1 GI:4702618

KEYWORDS EST.
SOURCE Dictyostelium discoidium
ORGANISM Dictyostelium discoidium
REFERENCE 1 (bases 1 to 171)
AUTHORS Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,
Yoshino, R., Mitra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,
Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.
TITLE Developmental cDNA in Dictyostelium discoidium
JOURNAL Unpublished (1998)
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664

FEATURES source

1. 171
/organism="Dictyostelium discoidium"
/mol_type="mRNA"
/strains="AX4"
/db_xref="taxon:44689"
/clone="SLK786"
/dev_stages="slug"
/clone_lib="Dictyostelium discoidium SL (H.Urushihara)"

ORIGIN

Alignment Scores:
Pred. No.: 78.6 Length: 171
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 55.70% Indels: 0
DB: 9 Gaps: 0

US-09-737-297-3 (1-16) x AU054137 (1-171)

Qy 7 AspValtyrGlnAsnIleGlnTy 14
Db 67 GATGTTATCAAAATATTCAATAC 44

RESULT 18 AU034493/c

LOCUS AU034493 Dictyostelium discoidium SL (H.Urushihara) Dictyostelium
DEFINITION AU034493 Dictyostelium discoidium SL (H.Urushihara) Dictyostelium
ACCESSION AU034493
VERSION AU034493.1 GI:3799917

KEYWORDS EST.
SOURCE Dictyostelium discoidium
ORGANISM Dictyostelium discoidium
REFERENCE 1 (bases 1 to 279)
AUTHORS Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,
Yoshino, R., Mitra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,
Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.
TITLE The Dictyostelium developmental cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development
JOURNAL DNA Res. 5 (6), 335-340 (1998)
MEDLINE 99156227
PUBMED 10048482
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp
PROJECT = 'Dictyostelium discoidium cDNA project in Japan'
POLYA=No.

FEATURES source

1. .279
/organism="Dictyostelium discoidium"
/mol_type="mRNA"
/strains="AX4"
/db_xref="taxon:44689"
/clone="SLC218"
/dev_stages="slug"
/clone_lib="Dictyostelium discoidium SL (H.Urushihara)"

ORIGIN

Alignment Scores:
Pred. No.: 152 Length: 279
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0

DEFINITION dc84g03.y1 NICHG_XGC_OO1 Xenopus laevis cDNA clone IMAGE:3403828 5' similar to TR:060635 "060635 TSPAN-1. [1] ;, mRNA sequence.
 ACCESSION BF047887
 VERSION BF047887.1 GI:10766390
 KEYWORDS EST.
 SOURCE Xenopus laevis (African clawed frog)
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae; Xenopus.
 REFERENCE 1 (bases 1 to 388)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Other ESTs: dc84g03.x1
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
 Trace considered overall poor quality
 Seq primer: -40RP from Gibco
 High quality sequence stop: 1.
 Location/Qualifiers
 1..388
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:3403828"
 /tissue_type="oocytes"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NICHG_XGC_OO1"
 /notes="vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.2 kb. Constructed by Life Technologies."
 ORIGIN
 Alignment Scores:
 Pred. No.: 237 Length: 388
 Score: 44.00 Matches: 7
 Percent Similarity: 76.92% Conservative: 3
 Best Local Similarity: 53.85% Mismatches: 3
 Query Match: 55.70% Indels: 0
 DB: 10 Gaps: 0
 US-09-737-297-3 (1-16) x BF047887 (1-388)
 Qy 2 GluglySerThr***AspValTyGlnAsnIleGlnTyr 14
 Db 73 GAGGGTACCCTCCAGAAATATCAGACGTTGATAT 35
 RESULT 28
 LOCUS BJ034056/c
 DEFINITION BJ034056 NIBB Mochii normalized Xenopus neurula library Xenopus laevis cDNA clone XL026c01 5', mRNA sequence.
 ACCESSION BJ034056
 KEYWORDS BJ034056.1 GI:17387747
 EST.
 SOURCE Xenopus laevis (African clawed frog)
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae; Xenopus.
 REFERENCE 1 (bases 1 to 415)
 AUTHORS Kitayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and Kohara, Y.
 TITLE Expressed genes in X. laevis embryo
 JOURNAL Unpublished (2001)

COMMENT Contact: Tadao Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshinigenes.nig.ac.jp
 The information of this clone is available through the following URL.
 http://xenopus.nibb.ac.jp.
 Location/Qualifiers
 1..415
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="XL026c01"
 /tissue_type="whole embryo"
 /dev_stage="stage 15"
 /clone_lib="NIBB Mochii normalized Xenopus neurula library"
 ORIGIN
 Alignment Scores:
 Pred. No.: 260 Length: 415
 Score: 44.00 Matches: 7
 Percent Similarity: 76.92% Conservative: 3
 Best Local Similarity: 53.85% Mismatches: 3
 Query Match: 55.70% Indels: 0
 DB: 12 Gaps: 0
 US-09-737-297-3 (1-16) x BJ034056 (1-415)
 Qy 2 GluglySerThr***AspValTyGlnAsnIleGlnTyr 14
 Db 97 GAGGGTACCCTCCAGAAATATCAGACGTTGATAT 59
 RESULT 29
 LOCUS CC469417
 DEFINITION CC469417 TV CHORI-240 Bos taurus genomic clone CH240_142A11, genomic survey sequence.
 ACCESSION CC469417
 KEYWORDS CC469417.1 GI:31655649
 SOURCE GSS.
 ORGANISM Bos taurus (cow)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
 REFERENCE 1 (bases 1 to 418)
 AUTHORS Costa, J.N., Mota, M. and Caetano, A.R.
 TITLE Brazil's Contribution to End-Sequencing the Bovine BAC Library CHORI-240
 JOURNAL Unpublished (2003)
 COMMENT Other GSSs: CH240_142A11.TJ
 Contact: Caetano AR
 Department of Biotechnology
 Embrapa Recursos Geneticos e Biotecnologia
 Parque Estacao Biologica, Final Av. W/5 Norte, Brasilia-DF C.P. 02372, 70770-900 Brasil
 Tel: 55 61 448 4778
 Fax: 55 61 340 3658
 Email: acaetano@cenargen.embrapa.br
 Clones are derived from the bovine BAC library CHORI-240 (http://www.chori.org/bacpac/bovine240.htm). Bases shown have Phred quality value equal to or higher than 20. Bases with quality value below 20 were masked with 'N'. For BAC library availability, please contact Pieter de Jong (pdejong@mail.choi.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/ordering_information.htm). This work was undertaken as part of the International Bovine BAC Mapping Consortium (IBBMC) by Embrapa Recursos Geneticos e Biotecnologia with financing from Conselho Nacional de Desenvolvimento Cientifico

e Tecnologico (CNPq), Brazil.
 Plate: 142 row: A column: 11
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 418.

FEATURES

Source
 Location/Qualifiers
 1..418
 /organism="Bos taurus"
 /mol_type="genomic DNA"
 /strain="bred: Hereford"
 /db_xref="taxon:9913"
 /clones="CH240_142A11"
 /sex="Male"
 /cell_type="Blood"
 /clone_lib="CHORI-240"
 /note="Vector: pRABAC1.3; Site 1: MboI; Site 2: MboI;
 Hereford bull LI Domino 99375; CHORI-240 Bovine BAC
 library (Male) produced by Pieter de Jong"

ORIGIN

Alignment Scores:
 Pred. No.: 262 Length: 418
 Score: 44.00 Matches: 7
 Percent Similarity: 76.92% Conservative: 3
 Best Local Similarity: 53.85% Mismatches: 3
 Query Match: 55.70% Indels: 0
 DB: 29 Gaps: 0

US-09-737-297-3 (1-16) x CC469417 (1-418)

QY 3 GlySerThr***AspValTyrGlnAsnIleGlnTyrAla 15
 ||||| :|||||
 DB 163 GGCAACCTGCATATATATATCAAAATCTCAACTATGCA 201

RESULT 30

BG037933/c
 LOCUS BG037933 438 bp mRNA linear EST 24-JAN-2001
 DEFINITION dc69hl1.y1 NICHD_XGC Embl Xenopus laevis cDNA clone IMAGE:3402500
 5' similar to TR:060635 O60635 TSPAN-1. [1] ;, mRNA sequence.

ACCESSION

VERSION BG037933

KEYWORDS

SOURCE BG037933.1 GI:12480518

ORGANISM

Xenopus laevis (African clawed frog)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 Xenopodinae; Xenopus.

REFERENCE

1 (bases 1 to 438)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

TITLE

Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Other_ESTs: dc69hl1.x1
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Washington University Genome Sequencing Center
 Clone distribution: Xenopus clones from this library are available
 through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
 Seq primer: -40RP from Gibco
 High quality sequence stop: 437.

FEATURES

source

Location/Qualifiers
 1..438
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:3402500"
 /tissue_type="embryo (stage 10)"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NICHD_XGC Embl"
 /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI;

ORIGIN

Alignment Scores:
 Pred. No.: 279 Length: 438
 Score: 44.00 Matches: 7
 Percent Similarity: 76.92% Conservative: 3
 Best Local Similarity: 53.85% Mismatches: 3
 Query Match: 55.70% Indels: 0
 DB: 10 Gaps: 0

US-09-737-297-3 (1-16) x BG037933 (1-438)

QY 2 GluGlySerThr***AspValTyrGlnAsnIleGlnTyr 14
 ||||| :|||||
 DB 115 GAGGTACCTCCAGAAATATCAGACGTTGAATAT 77

RESULT 31

C83991/c
 LOCUS C83991 456 bp mRNA linear EST 28-APR-1999
 DEFINITION C83991 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
 discoideum cDNA clone SLK667, mRNA sequence.

ACCESSION

VERSION C83991.1 GI:2723597

KEYWORDS

SOURCE Dictyostelium discoideum

ORGANISM

Dictyostelium discoideum
 Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

REFERENCE

1 (bases 1 to 456)
 Morio, T., Urushihara, H., Saito, T., Uga, Y., Mizuno, H., Yoshida, M.,
 Yoshino, R., Mitra, B. N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,
 Williams, J., Maeda, M., Takeuchi, I., Ohlali, H. and Tanaka, Y.

TITLE

The Dictyostelium developmental cDNA project: generation and
 analysis of expressed sequence tags from the first-finger stage of
 development

JOURNAL

DNA Res. 5 (6), 335-340 (1998)

MEDLINE

99156227

COMMENT

10048482
 Contact: Hideko Urushihara
 Institute of Biological Sciences
 University of Tsukuba
 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
 Tel: 81-298-53-4664
 Fax: 81-298-53-6614
 Email: hideko@biol.tsukuba.ac.jp

PROJECT = 'Dictyostelium discoideum cDNA project in Japan'

POLYA-No.

FEATURES

Location/Qualifiers

1..456
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="SLK667"
 /dev_stage="slug"
 /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"

ORIGIN

Alignment Scores:
 Pred. No.: 295 Length: 456
 Score: 44.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 55.70% Indels: 0
 DB: 13 Gaps: 0

US-09-737-297-3 (1-16) x C83991 (1-456)

QY 7 AspValTyrGlnAsnIleGlnTyr 14

DB 400 GATGTTTATCAAAATATCAATAC 377

Cloned unidirectionally. Primer: Oligo dt. Average insert
 size 1.55 kb. Constructed by Life Technologies. Note: This
 is a Xenopus gene Collection (XGC) library."

RESULT 32
 BG408335/c
 LOCUS
 DEFINITION dc9909.y1 NICHDXGC 001 Xenopus laevis cDNA clone IMAGE:3405496 5', similar to TR:O60635 TSPAN-1. [1] ;, mRNA sequence.
 ACCESSION BG408335
 VERSION BG408335.1 GI:13314684
 KEYWORDS EST.
 SOURCE Xenopus laevis (African clawed frog)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus.
 1 (bases 1 to 457)
 REFERENCE NC1-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 TITLE Unpublished (1997)
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps@mail.nih.gov
 Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
 Seq primer: -40RP from Gibco
 High quality sequence stop: 413.
 FEATURES
 source
 1. .457
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:3405496"
 /tissue_type="oocytes"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NICHDXGC 001"
 /notes="Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.2 kb. Constructed by Life Technologies."
 ORIGIN
 Alignment Scores:
 Pred. No.: 296 Length: 457
 Score: 44.00 Matches: 7
 Percent Similarity: 76.92% Conservatative: 3
 Best Local Similarity: 53.85% Mismatches: 3
 Query Match: 55.70% Indels: 0
 DB: 12 Gaps: 0
 US-09-737-297-3 (1-16) x BG408335 (1-457)
 QY 2 GluGlySerThr***AspValTyGlnAsnIleGlnTyr 14
 |||||:|||||
 DB 136 GAGGGTACCCTCCAGAAATATCAGACGTTGAATAT 98
 RESULT 33
 BJ070599/c
 LOCUS
 DEFINITION BJ070599 NIBB Mochii normalized Xenopus tailbud library Xenopus laevis cDNA clone X1091b02 5', mRNA sequence.
 ACCESSION BJ070599
 VERSION BJ070599.1 GI:17500788
 KEYWORDS EST.
 SOURCE Xenopus laevis (African clawed frog)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus.
 1 (bases 1 to 477)
 REFERENCE Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara,Y.

Expressed genes in X. laevis embryo
 Unpublished (2001)
 Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshinigenes.nig.ac.jp
 The information of this clone is available through the following URL.
<http://xenopus.nibb.ac.jp>.
 FEATURES
 source
 1. .477
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="X1091b02"
 /tissue_type="whole embryo"
 /dev_stage="stage 25"
 /clone_lib="NIBB Mochii normalized Xenopus tailbud library"
 ORIGIN
 Alignment Scores:
 Pred. No.: 313 Length: 477
 Score: 44.00 Matches: 7
 Percent Similarity: 76.92% Conservatative: 3
 Best Local Similarity: 53.85% Mismatches: 3
 Query Match: 55.70% Indels: 0
 DB: 12 Gaps: 0
 US-09-737-297-3 (1-16) x BJ070599 (1-477)
 QY 2 GluGlySerThr***AspValTyGlnAsnIleGlnTyr 14
 |||||:|||||
 DB 75 GAGGGTACCCTCCAGAAATATCAGACGTTGAATAT 37
 RESULT 34
 BF426652/c
 LOCUS
 DEFINITION BF426652 481 bp mRNA linear EST 19-FEB-2003
 laevis cDNA clone IMAGE:3744801 5', similar to TR:O60635 O60635 TSPAN-1. [1] ;, mRNA sequence.
 ACCESSION BF426652
 VERSION BF426652.1 GI:11438101
 KEYWORDS EST.
 SOURCE Xenopus laevis (African clawed frog)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus.
 1 (bases 1 to 481)
 REFERENCE Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterston,R. and Wilson,R.
 WashU Xenopus EST project, 1999
 Unpublished (1999)
 Contact: Sandy Clifton, Ph.D.
 WashU Xenopus EST project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Library constructed by Bruce Blumberg
 Library normalized by Jihwan Song
 DNA Sequencing by: Washington University Genome Sequencing Center
 Source lab clone id - xinneg006f10 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (image@image.llnl.gov) for further information.
 Seq primer: -40RP from Gibco


```

/dev_stage="slug"
/clone_lib="Dictyostelium discoideum SL (H.Urushihara)"

ORIGIN
Alignment Scores:
Pred. No.: 324 Length: 489
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 55.70% Indels: 0
DB: 9 Gaps: 0

US-09-737-297-3 (1-16) x AU054046 (1-489)

Qy 7 AspValTyGlnAsnIleGlnTyR 14
|||||
Db 435 GATGTTTATCAAAATATTCATAC 412

RESULT 37
AU034593/c
LOCUS AU034593 518 bp mRNA linear EST 28-APR-1999
DEFINITION AU034593 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
discoideum cDNA clone SLC585, mRNA sequence.
ACCESSION AU034593
VERSION AU034593.1 GI:3800017
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 518)
AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
TITLE The Dictyostelium developmental cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development
JOURNAL DNA Res. 5 (6), 335-340 (1998)
MEDLINE 99156227
PUBMED 10048482
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp
PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
POLYA=No. Location/Qualifiers
source
1. 518
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SLC585"
/dev_stage="slug"
/clone_lib="Dictyostelium discoideum SL (H.Urushihara)"

ORIGIN
Alignment Scores:
Pred. No.: 350 Length: 518
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 55.70% Indels: 0
DB: 9 Gaps: 0

US-09-737-297-3 (1-16) x AU034593 (1-518)

Qy 7 AspValTyGlnAsnIleGlnTyR 14
|||||
Db 464 GATGTTTATCAAAATATTCATAC 441

```

```

RESULT 38
BZ991276
LOCUS BZ991276 519 bp DNA linear GSS 25-MAR-2003
DEFINITION PUD0542TD ZM_0.6_1.0 KB Zea mays genomic clone ZMMBTal64G11,
genomic survey sequence.
ACCESSION BZ991276
VERSION BZ991276.1 GI:29232913
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 519)
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
Maize Genomics Consortium
Unpublished (2003)
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5943
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.
Location/Qualifiers
source
1. 519
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTal64G11"
/clone_lib="ZM_0.6_1.0 KB"
/notes="Vector: pCR4-TOPO; Site:1: EcoRI; 0.6-1.0 kb high
Cot selected genomic DNA library"

ORIGIN
Alignment Scores:
Pred. No.: 351 Length: 519
Score: 44.00 Matches: 8
Percent Similarity: 76.92% Conservative: 2
Best Local Similarity: 61.54% Mismatches: 3
Query Match: 55.70% Indels: 0
DB: 28 Gaps: 0

US-09-737-297-3 (1-16) x BZ991276 (1-519)

Qy 2 GluGlySerThr***AspValTyGlnAsnIleGlnTyR 14
|||||
Db 106 GAAGGATCAGCATGGAAGCATATCAGATCCCAATTC 144

RESULT 39
AQ405209/c
LOCUS AQ405209 533 bp DNA linear GSS 13-MAR-1999
DEFINITION HS_5038_B2_E11_T7 RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=614 Col=22 Row=J, genomic survey sequence.
ACCESSION AQ405209
VERSION AQ405209.1 GI:4411017
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 533)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
MEDLINE

```


10449764
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@jg.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: <http://www.hsc.washington.edu>
Plate: 614 Row: J Column: 22
Seq primer: 17
Class: BAC ends
High quality sequence stop: 533.
Location/Qualifiers
1. 533
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=614 Col=22 Row=J"
/sex="male"
/clone_lib="RPCI-11 Human Male BAC Library"
/note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

Alignment Scores:
Pred. No.: 364 Length: 533
Score: 44.00 Matches: 8
Percent Similarity: 78.57% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 3
Query Match: 55.70% Indels: 0
DB: 28 Gaps: 0

US-09-737-297-3 (1-16) x A0405209 (1-533)

QY 2 GIUGlySerThr**AspValTyGlnAsnIleGlnTyAla 15
DB 212 GATGTTCCACGCGTGAATTCACCAAGATTTCAGTATAGT 171

RESULT 40
C91372/c
LOCUS
DEFINITION
Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
discoideum cDNA clone SSK136, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M., Yoshino, R., Mitra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H., Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.
Developmental cDNA in Dictyostelium discoideum
Unpublished (1998)
Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp.
Location/Qualifiers
1. 535

Alignment Scores:
Pred. No.: 364 Length: 533
Score: 44.00 Matches: 7
Percent Similarity: 78.57% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 3
Query Match: 55.70% Indels: 0
DB: 28 Gaps: 0

US-09-737-297-3 (1-16) x A0405209 (1-533)

QY 2 GIUGlySerThr**AspValTyGlnAsnIleGlnTyAla 15
DB 212 GATGTTCCACGCGTGAATTCACCAAGATTTCAGTATAGT 171

RESULT 40
C91372/c
LOCUS
DEFINITION
Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
discoideum cDNA clone SSK136, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M., Yoshino, R., Mitra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H., Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.
Developmental cDNA in Dictyostelium discoideum
Unpublished (1998)
Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp.
Location/Qualifiers
1. 535

Alignment Scores:
Pred. No.: 366 Length: 535
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 55.70% Indels: 0
DB: 13 Gaps: 0

US-09-737-297-3 (1-16) x C91372 (1-535)

QY 7 AspValTyGlnAsnIleGlnTy 14
DB 351 GATGTTTATCAAAATATTCAATAC 328

RESULT 41
BM879901/c
LOCUS
DEFINITION
Strongyloides ratti PA female naive PAMP1 v1
Strongyloides ratti cDNA 5', similar to TR:Q9XX09 Q9XX09 Y39A1A.15C
PROTEIN [1] ; mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, S., Bennett, J., Franklin, C., Tsagaris, V., R., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
The Washington Univ. Nematode EST Project, 1999
Unpublished (1999)
Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estw@wustl.edu
The library was constructed by Brandi Chiapelli and Dr. James McCarter at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dt priming (Dyna). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis Kit from Clontech. Directionally cloned into the UDG sites of PAMP1. Dissected nematode tissues were provided by Dr. Alan Scott (ascott@hph.edu) of the School of Public Hygiene and Public Health at John Hopkins University in Baltimore, MD.
High quality sequence stop: 395.
Location/Qualifiers
1. 539

Alignment Scores:
Pred. No.: 366 Length: 535
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 55.70% Indels: 0
DB: 13 Gaps: 0

US-09-737-297-3 (1-16) x C91372 (1-535)

QY 7 AspValTyGlnAsnIleGlnTy 14
DB 351 GATGTTTATCAAAATATTCAATAC 328

RESULT 41
BM879901/c
LOCUS
DEFINITION
Strongyloides ratti PA female naive PAMP1 v1
Strongyloides ratti cDNA 5', similar to TR:Q9XX09 Q9XX09 Y39A1A.15C
PROTEIN [1] ; mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, S., Bennett, J., Franklin, C., Tsagaris, V., R., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
The Washington Univ. Nematode EST Project, 1999
Unpublished (1999)
Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estw@wustl.edu
The library was constructed by Brandi Chiapelli and Dr. James McCarter at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dt priming (Dyna). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis Kit from Clontech. Directionally cloned into the UDG sites of PAMP1. Dissected nematode tissues were provided by Dr. Alan Scott (ascott@hph.edu) of the School of Public Hygiene and Public Health at John Hopkins University in Baltimore, MD.
High quality sequence stop: 395.
Location/Qualifiers
1. 539

St. Louis. The cDNA was made by using Dynabead oligo-dT priming (Dyna). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis Kit from Clontech. Directionally cloned into the UDG sites of pAMP1. Parasitic adult females were collected from naive animals and provided by Dr. Mark Viney of Bristol, UK."

ORIGIN

Alignment Scores:
 Pred. No.: 369 Length: 539
 Score: 44.00 Matches: 8
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 55.70% Indels: 0
 DB: 12 Gaps: 0

US-09-737-297-3 (1-16) x BM879901 (1-539)

QY 7 AspValTyrGlnAsnIleGlnTyr 14

Db 456 GATGTATACCAAGACATTCAATAC 433

RESULT 42

LOCUS BG226255/c

DEFINITION kq19d02.v1.TB95TW-SSR Strongyloides stercoralis

VERSION WP:139A1A.15C CE20232 ; mRNA sequence.

SOURCE BG226255

ORGANISM Strongyloides stercoralis

REFERENCE BG226255.1 GI:12713810

AUTHORS EST.

KEYWORDS Strongyloides stercoralis

ORGANISM Strongyloides stercoralis

REFERENCE BG226255.1 GI:12713810

AUTHORS EST.

KEYWORDS Strongyloides stercoralis

ORGANISM Strongyloides stercoralis

REFERENCE BG226255.1 GI:12713810

AUTHORS EST.

KEYWORDS Strongyloides stercoralis

ORGANISM Strongyloides stercoralis

REFERENCE BG226255.1 GI:12713810

AUTHORS EST.

KEYWORDS Strongyloides stercoralis

ORGANISM Strongyloides stercoralis

REFERENCE BG226255.1 GI:12713810

AUTHORS EST.

KEYWORDS Strongyloides stercoralis

ORGANISM Strongyloides stercoralis

REFERENCE BG226255.1 GI:12713810

AUTHORS EST.

KEYWORDS Strongyloides stercoralis

ORGANISM Strongyloides stercoralis

REFERENCE BG226255.1 GI:12713810

AUTHORS EST.

KEYWORDS Strongyloides stercoralis

ORGANISM Strongyloides stercoralis

REFERENCE BG226255.1 GI:12713810

AUTHORS EST.

KEYWORDS Strongyloides stercoralis

ORGANISM Strongyloides stercoralis

REFERENCE BG226255.1 GI:12713810

AUTHORS EST.

KEYWORDS Strongyloides stercoralis

ORGANISM Strongyloides stercoralis

REFERENCE BG226255.1 GI:12713810

AUTHORS EST.

KEYWORDS Strongyloides stercoralis

ORGANISM Strongyloides stercoralis

REFERENCE BG226255.1 GI:12713810

AUTHORS EST.

KEYWORDS Strongyloides stercoralis

ORGANISM Strongyloides stercoralis

REFERENCE BG226255.1 GI:12713810

AUTHORS EST.

KEYWORDS Strongyloides stercoralis

ORGANISM Strongyloides stercoralis

REFERENCE BG226255.1 GI:12713810

AUTHORS EST.

KEYWORDS Strongyloides stercoralis

ORGANISM Strongyloides stercoralis

REFERENCE BG226255.1 GI:12713810

ORIGIN

Alignment Scores:
 Pred. No.: 371 Length: 541
 Score: 44.00 Matches: 8
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 55.70% Indels: 0
 DB: 12 Gaps: 0

US-09-737-297-3 (1-16) x BG226255 (1-541)

QY 7 AspValTyrGlnAsnIleGlnTyr 14

Db 425 GATGTATACCAAGACATTCAATAC 402

RESULT 43

LOCUS AU053352/c

DEFINITION AU053352 Dictyostelium discoideum SL (H. Urushihara)

VERSION AU053352

SOURCE AU053352.1 GI:4701834

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

(range, 100-1700) ."

ORIGIN

Alignment Scores:
 Pred. No.: 371 Length: 541
 Score: 44.00 Matches: 8
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 55.70% Indels: 0
 DB: 12 Gaps: 0

US-09-737-297-3 (1-16) x BG226255 (1-541)

QY 7 AspValTyrGlnAsnIleGlnTyr 14

Db 425 GATGTATACCAAGACATTCAATAC 402

RESULT 43

LOCUS AU053352/c

DEFINITION AU053352 Dictyostelium discoideum SL (H. Urushihara)

VERSION AU053352

SOURCE AU053352.1 GI:4701834

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

(range, 100-1700) ."

ORIGIN

Alignment Scores:
 Pred. No.: 371 Length: 541
 Score: 44.00 Matches: 8
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 55.70% Indels: 0
 DB: 12 Gaps: 0

US-09-737-297-3 (1-16) x BG226255 (1-541)

QY 7 AspValTyrGlnAsnIleGlnTyr 14

Db 425 GATGTATACCAAGACATTCAATAC 402

RESULT 43

LOCUS AU053352/c

DEFINITION AU053352 Dictyostelium discoideum SL (H. Urushihara)

VERSION AU053352

SOURCE AU053352.1 GI:4701834

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE AU053352

AUTHORS EST.

KEYWORDS Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

TITLE The WashU-HMMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMMI Mouse EST Project
Washington University School
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mousestewatson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:527404
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 256.
Location/Qualifiers
1. 569
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6 x DBA"
/db_xref="taxon:10090"
/clone="IMAGE:906740"
/sex="pooled"
/tissue_type="embryo"
/dev_stage="7.5dpc"
/lab_host="DH12S"
/clone_lib="Bedington mouse embryonic region"
/note="Organ: whole embryo; Vector: pCMV.SPORT; Site:1:
SalI; Site_2: NotI; Cloned unidirectionally. Primer:
Oligo dr. Gastrulating embryos were collected at 7.5dpc
from C57BL/6 x DBA matings, excluding embryos that had
developed head folds and all extraembryonic tissues.
Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).
Referenced in Development 121, 2479-2489 (1995)"

ORIGIN
Alignment Scores:
Pred. No.: 397 Length: 569
Score: 44.00 Matches: 8
Percent Similarity: 71.43% Conservative: 2
Best Local Similarity: 57.14% Mismatches: 4
Query Match: 55.70% Indels: 0
DB: 9 Gaps: 0

US-09-737-297-3 (1-16) x AA522323 (1-569)

QY 3 GlySerThr***AspValTyGlnAsnIleGlnTyAlaGly 16
DB 318 GGCAGACACGCTGGAGTTATCAAAACTCCGATATTCAGG 359

RESULT 49
C94080/c
LOCUS C94080 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
discoideum cDNA clone SSG382, mRNA sequence.
ACCESSION C94080
VERSION C94080.1 GI:3218695
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
1 (bases 1 to 569)
REFERENCE Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,
AUTHORS Yoshino, R., Mitra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,
Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.
Developmental cDNA in Dictyostelium discoideum
Unpublished (1998)
Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp

PROJECT = Dictyostelium discoideum cDNA project in Japan.

FEATURES

source
1..569
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SSG382"
/dev_stage="slug"
/clone_lib="Dictyostelium discoideum SS (H.Urushihara)"

ORIGIN

Alignment Scores:
Pred. No.: 397 Length: 569
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 55.70% Indels: 0
DB: 13 Gaps: 0

US-09-737-297-3 (1-16) x C94080 (1-569)

QY 7 AspValtyrGlnAsnIleGlnTyr 14

DB 464 GATGTTTATCAAAATATTCAATAC 441

RESULT 50

C93278/c

LOCUS

C93278 607 bp mRNA linear EST 12-JUL-1999
C93278 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
discoideum cDNA clone SSM516, mRNA sequence.

ACCESSION

C93278

VERSION

C93278.1 GI:3075154

KEYWORDS

EST.

SOURCE

Dictyostelium discoideum

ORGANISM

Dictyostelium discoideum

REFERENCE

1 (bases 1 to 607)

AUTHORS

Morita, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,
Yoshino, R., Mitra, B. N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,
Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.

TITLE

The Dictyostelium developmental cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development

JOURNAL

DNA Res. 5 (6), 335-340 (1998)

MEDLINE

99156227

PUBMED

10048482

COMMENT

Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp
PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
POLYA=No.

FEATURES

source
1..607
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SSM516"
/dev_stage="slug"
/clone_lib="Dictyostelium discoideum SS (H.Urushihara)"

ORIGIN

Alignment Scores:
Pred. No.: 434 Length: 607
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 55.70% Indels: 0

DB: 13 Gaps: 0

US-09-737-297-3 (1-16) x C93278 (1-607)

QY 7 AspValtyrGlnAsnIleGlnTyr 14

DB 424 GATGTTTATCAAAATATTCAATAC 401

Search completed: March 1, 2004, 11:33:46
Job time : 2514 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

DM protein - nucleic search, using frame_plus_p2n model

run on: March 1, 2004, 10:16:24 ; Search time 78 Seconds
(without alignments)
113.836 Million cell updates/sec

Title: US-09-737-297-3

Perfect score: 79

Sequence: 1 ARGSTXDYQNIQVAG 16

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Command line parameters:
-MODEL=frame+P2n.model -DEV=xlp
-Q/cgn2_1/USPFO_Spool_P/US09737297/runat_01032004_085324_3487/app_query.fasta_1.199
-DB=Issued Patents NA -QMT=fastcap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=150 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=50
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09737297 -CGN 1 85 -runat_01032004_085324_3487 -NCPUS=6 -ICPU=3
-NO MAP -LARGOQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREAS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6C.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	53.2	5329	4	US-09-562-702A-19
2	42	53.2	5329	4	US-09-561-818A-19
3	42	53.2	5689	4	US-09-562-702A-17
4	42	53.2	5689	4	US-09-561-818A-17
5	42	53.2	11384	4	US-08-961-527-45
6	41	51.9	675	1	US-07-744-570B-1
7	41	51.9	43360	4	US-09-453-702B-206
8	41	51.9	43325	4	US-09-453-702B-261
9	41	51.9	1830121	4	US-09-557-884-1
10	41	51.9	1830121	4	US-09-643-990A-1
11	40	50.6	447	4	US-09-621-976-13424
12	40	50.6	1812	4	US-09-732-615-28

Sequence 235, Appl	US-08-961-527-235	1766	4	Sequence 2, Appl
Sequence 13, Appl	US-09-551-826D-13	939	4	Sequence 2, Appl
Sequence 5, Appl	US-07-921-796-5	2466	1	Sequence 1, Appl
Sequence 7, Appl	US-07-921-796-7	3089	1	Sequence 1, Appl
Sequence 2620, Ap	US-09-328-352-2620	3783	4	Sequence 2, Appl
Sequence 4595, Ap	US-08-005-180A-2	3246	3	Sequence 4, Appl
Sequence 4399, Ap	US-08-252-951A-4595	441	4	Sequence 4, Appl
Sequence 221, App	US-09-252-951A-4399	876	4	Sequence 2, Appl
Sequence 2901, Ap	US-09-328-352-2901	1098	4	Sequence 2, Appl
Sequence 1177, Ap	US-09-134-000C-1177	1368	4	Sequence 1, Appl
Sequence 3818, Ap	US-09-328-352-3818	1632	1	Sequence 6, Appl
Sequence 6, Appl	US-08-036-210-6	2173	1	Sequence 6, Appl
Sequence 6, Appl	US-08-443-609-6	2173	2	Sequence 6, Appl
Sequence 10, Appl	US-09-361-096A-6	2173	4	Sequence 10, Appl
Sequence 10, Appl	US-08-036-210-10	2309	1	Sequence 10, Appl
Sequence 10, Appl	US-08-449-609-10	2309	2	Sequence 10, Appl
Sequence 14, Appl	US-09-361-096A-10	2309	4	Sequence 14, Appl
Sequence 14, Appl	US-08-036-210-14	2692	1	Sequence 14, Appl
Sequence 14, Appl	US-08-443-609-14	2692	2	Sequence 14, Appl
Sequence 14, Appl	US-09-361-096A-14	2692	4	Sequence 14, Appl
Sequence 181, App	US-09-107-532A-681	2784	4	Sequence 181, App
Sequence 21, Appl	US-08-936-165A-194	3191	4	Sequence 21, Appl
Sequence 21, Appl	US-08-036-210-21	3973	1	Sequence 21, Appl
Sequence 21, Appl	US-08-449-609-21	3973	2	Sequence 21, Appl
Sequence 234, App	US-09-361-096A-21	3973	4	Sequence 234, App
Sequence 111, App	US-08-956-171E-234	4721	4	Sequence 111, App
Sequence 2, Appl	US-09-453-702B-111	2119	4	Sequence 2, Appl
Sequence 2, Appl	US-09-103-840A-2	4403765	3	Sequence 2, Appl
Sequence 1, Appl	US-09-103-840A-1	4403765	3	Sequence 1, Appl
Sequence 1, Appl	US-09-103-840A-1	4411529	3	Sequence 1, Appl
Sequence 46, Appl	US-09-189-060B-46	234	3	Sequence 46, Appl
Sequence 46, Appl	US-09-189-060B-54	234	3	Sequence 46, Appl
Sequence 2067, Ap	US-09-543-681A-2067	240	4	Sequence 2067, Ap
Sequence 18487, A	US-09-711-184-4	382	4	Sequence 18487, A
Sequence 26, Appl	US-09-621-976-18487	493	4	Sequence 26, Appl
Sequence 34, Appl	US-08-275-526C-34	600	3	Sequence 34, Appl
Sequence 34, Appl	US-08-275-526C-34	600	3	Sequence 34, Appl
Sequence 26, Appl	US-09-076-677-26	600	4	Sequence 26, Appl
Sequence 26, Appl	US-09-076-677-34	600	4	Sequence 26, Appl
Sequence 26, Appl	US-09-073-055-26	600	4	Sequence 26, Appl
Sequence 30, Appl	US-09-073-055-30	600	4	Sequence 30, Appl
Sequence 30, Appl	US-08-275-526C-30	681	3	Sequence 30, Appl
Sequence 32, Appl	US-09-076-677-30	681	4	Sequence 32, Appl
Sequence 32, Appl	US-09-076-677-32	681	4	Sequence 32, Appl
Sequence 32, Appl	US-09-073-055-30	681	4	Sequence 32, Appl
Sequence 32, Appl	US-09-073-055-32	681	4	Sequence 32, Appl
Sequence 13272, A	US-09-252-951A-13272	732	4	Sequence 13272, A
Sequence 1, Appl	US-08-275-526C-1	1022	3	Sequence 1, Appl
Sequence 1, Appl	US-08-275-526C-35	1022	3	Sequence 1, Appl
Sequence 1, Appl	US-09-076-677-1	1022	4	Sequence 1, Appl
Sequence 35, Appl	US-09-076-677-35	1022	4	Sequence 35, Appl
Sequence 35, Appl	US-09-073-055-1	1022	4	Sequence 35, Appl
Sequence 35, Appl	US-09-073-055-35	1022	4	Sequence 35, Appl
Sequence 13107, A	US-09-252-951A-13107	1056	4	Sequence 13107, A
Sequence 2, Appl	US-08-575-964-2	1207	1	Sequence 2, Appl
Sequence 2, Appl	US-08-963-500-2	1207	2	Sequence 2, Appl
Sequence 129, App	US-09-711-164-129	2187	4	Sequence 129, App
Sequence 4, Appl	US-07-882-925A-4	2188	1	Sequence 4, Appl
Sequence 3, Appl	US-08-184-012C-4	4713	3	Sequence 3, Appl
Sequence 7, Appl	US-09-068-655-3	4716	2	Sequence 7, Appl
Sequence 5, Appl	US-09-011-745-5	6312	3	Sequence 5, Appl
Sequence 5, Appl	US-07-882-925A-5	6751	1	Sequence 5, Appl
Sequence 1, Appl	US-08-184-012C-1	6751	1	Sequence 1, Appl
Sequence 1, Appl	US-09-315-127-4	8088	4	Sequence 1, Appl
Sequence 1, Appl	US-08-716-351A-1	8535	3	Sequence 1, Appl
Sequence 3, Appl	US-09-026-039-3	11721	4	Sequence 3, Appl
Sequence 17, Appl	US-09-028-039-3	11721	4	Sequence 17, Appl
Sequence 1, Appl	US-09-128-155-17	176373	3	Sequence 1, Appl
Sequence 1, Appl	US-09-198-452A-1	1230025	4	Sequence 1, Appl

86	36.5	46.2	624	3	US-08-998-416-1151	Sequence 1151, Ap
87	36	45.6	240	3	US-09-189-060B-52	Sequence 52, Appl
88	36	45.6	265	4	US-09-313-294A-1862	Sequence 1862, Ap
89	36	45.6	345	1	US-08-171-385-20	Sequence 20, Appl
90	36	45.6	345	1	US-08-361-441B-20	Sequence 20, Appl
91	36	45.6	366	4	US-09-134-000C-1026	Sequence 1026, Ap
92	36	45.6	396	4	US-09-621-976-8159	Sequence 8159, Ap
93	36	45.6	467	4	US-09-621-976-8159	Sequence 1857, A
94	36	45.6	534	4	US-09-621-976-8159	Sequence 3266, Ap
95	36	45.6	768	4	US-09-134-000C-676	Sequence 676, App
96	36	45.6	771	3	US-08-991-789A-241	Sequence 241, App
97	36	45.6	771	4	US-09-063-451-241	Sequence 241, App
98	36	45.6	771	4	US-09-598-326-241	Sequence 241, App
99	36	45.6	771	4	US-09-289-198-241	Sequence 241, App
100	36	45.6	771	4	US-09-428-725-241	Sequence 241, App
101	36	45.6	891	4	US-09-328-352-456	Sequence 456, App
102	36	45.6	1215	4	US-09-328-352-2728	Sequence 2728, Ap
103	36	45.6	1272	4	US-09-134-001C-762	Sequence 762, App
104	36	45.6	1296	4	US-09-328-352-3266	Sequence 3266, Ap
105	36	45.6	1365	2	US-08-870-827-4	Sequence 4, Appli
106	36	45.6	1365	4	US-09-317-179-4	Sequence 4, Appli
107	36	45.6	1455	4	US-09-328-352-723	Sequence 723, App
108	36	45.6	2408	2	US-08-870-827-5	Sequence 5, Appli
109	36	45.6	2408	4	US-09-317-179-5	Sequence 5, Appli
110	36	45.6	2592	4	US-09-489-039A-3008	Sequence 3008, Ap
111	36	45.6	2685	4	US-09-543-681A-552	Sequence 552, App
112	36	45.6	3589	4	US-09-221-017B-265	Sequence 265, App
113	36	45.6	4845	4	US-09-926-171B-265	Sequence 1011, Ap
114	36	45.6	4892	4	US-09-976-594-167	Sequence 167, App
115	36	45.6	5158	4	US-09-023-655-1347	Sequence 1347, Ap
116	36	45.6	5198	1	US-08-123-761A-1	Sequence 1, Appli
117	36	45.6	5433	4	US-09-562-702A-15	Sequence 15, Appl
118	36	45.6	5433	4	US-09-561-818A-15	Sequence 15, Appl
119	36	45.6	5613	4	US-09-562-702A-13	Sequence 13, Appl
120	36	45.6	5613	4	US-09-561-818A-13	Sequence 13, Appl
121	36	45.6	5813	4	US-09-561-709B-10	Sequence 10, Appl
122	36	45.6	8060	3	US-08-766-528-1	Sequence 1, Appli
123	36	45.6	8132	3	US-08-766-528-3	Sequence 3, Appli
124	36	45.6	9636	1	US-08-323-170B-1	Sequence 1, Appli
125	36	45.6	9636	4	US-08-954-441-1	Sequence 1, Appli
126	36	45.6	12720	1	US-08-403-866-11	Sequence 11, Appl
127	36	45.6	38653	4	US-09-922-445-1	Sequence 1, Appli
128	35.5	44.9	8557	3	US-08-998-416-78	Sequence 78, Appl
129	35	44.3	257	4	US-09-134-000C-2054	Sequence 2054, Ap
130	35	44.3	297	4	US-09-252-991A-12192	Sequence 12192, A
131	35	44.3	476	4	US-09-551-169A-38	Sequence 38, Appl
132	35	44.3	555	1	US-08-088-633-5	Sequence 5, Appli
133	35	44.3	555	1	US-08-445-756-5	Sequence 5, Appli
134	35	44.3	555	2	US-08-441-750-5	Sequence 5, Appli
135	35	44.3	555	2	PCT-US92-02521-5	Sequence 5, Appli
136	35	44.3	572	1	US-08-044-621D-2	Sequence 2, Appli
137	35	44.3	572	1	US-08-044-621D-3	Sequence 3, Appli
138	35	44.3	579	1	US-08-709-912-19	Sequence 19, Appl
139	35	44.3	579	2	US-09-047-370-19	Sequence 19, Appl
140	35	44.3	642	1	US-08-104-445-2	Sequence 2, Appli
141	35	44.3	700	1	US-08-044-621D-1	Sequence 1, Appli
142	35	44.3	1033	4	US-09-252-991A-6421	Sequence 6421, Ap
143	35	44.3	1036	4	US-09-328-352-2030	Sequence 2030, Ap
144	35	44.3	1077	4	US-09-252-991A-10970	Sequence 10970, A
145	35	44.3	1176	4	US-09-489-039A-5553	Sequence 5553, Ap
146	35	44.3	1179	4	US-09-252-991A-11273	Sequence 11273, A
147	35	44.3	1293	4	US-09-252-991A-6045	Sequence 6045, Ap
148	35	44.3	1349	4	US-09-570-856B-2	Sequence 2, Appli
149	35	44.3	1375	2	US-08-468-812-1	Sequence 1, Appli
150	35	44.3				

GENERAL INFORMATION:

APPLICANT: Yurchenco, Peter

TITLE OF INVENTION: Laminin 2 and Methods for Its Use

FILE REFERENCE: 99-274-B

CURRENT APPLICATION NUMBER: US/09/562,702A

CURRENT FILING DATE: 2000-04-28

PRIOR APPLICATION NUMBER: 60/155,945

PRIOR FILING DATE: 1999-09-24

PRIOR APPLICATION NUMBER: 60/143,289

PRIOR FILING DATE: 1999-07-12

PRIOR APPLICATION NUMBER: 60/139,198

PRIOR FILING DATE: 1999-06-15

PRIOR APPLICATION NUMBER: 60/131,720

PRIOR FILING DATE: 1999-04-30

NUMBER OF SEQ ID NOS: 32

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 19

LENGTH: 5329

TYPE: DNA

ORGANISM: Mus musculus

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(5175)

US-09-562-702A-19

Alignment Scores:

Pred. No.: 117

Score: 42.00

Percent Similarity: 71.43%

Best Local Similarity: 50.00%

Query Match: 53.16%

DB: 4

Length: 5329

Matches: 7

Conservative: 3

Mismatches: 4

Indels: 0

Gaps: 0

US-09-737-297-3 (1-16) x US-09-562-702A-19 (1-5329)

OY 3 GlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16

DB 3723 GCGAGACAGCTGGAGTTCATCAAAACCTCCGATATTCAGGG 3764

RESULT 2

US-09-561-818A-19

Sequence 19, Application US/09561818A

Patent No. 6638907

GENERAL INFORMATION:

APPLICANT: Kortessmaa, Jarrko

APPLICANT: Tryggvason, Karl

TITLE OF INVENTION: Laminin 8 and Methods For Its Use

FILE REFERENCE: 99/274-D

CURRENT APPLICATION NUMBER: US/09/561,818A

CURRENT FILING DATE: 2000-04-28

NUMBER OF SEQ ID NOS: 28

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 19

LENGTH: 5329

TYPE: DNA

ORGANISM: Mus musculus

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(5175)

US-09-561-818A-19

Alignment Scores:

Pred. No.: 117

Score: 42.00

Percent Similarity: 71.43%

Best Local Similarity: 50.00%

Query Match: 53.16%

DB: 4

Length: 5329

Matches: 7

Conservative: 3

Mismatches: 4

Indels: 0

Gaps: 0

US-09-737-297-3 (1-16) x US-09-561-818A-19 (1-5329)

OY 3 GlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16

DB 3723 GCGAGACAGCTGGAGTTCATCAAAACCTCCGATATTCAGGG 3764

ALIGNMENTS

RESULT 1

US-09-562-702A-19

Sequence 19, Application US/09562702A

Patent No. 6632790


```

Db      3723  GGCAGACAGCTGGAGTTTATCAAAAACCTCCGATATTCAGGG 3766
RESULT 3
JS-09-562-702A-17
; Sequence 17, Application US/09562702A
; Patent No. 663290
; GENERAL INFORMATION:
; APPLICANT: Yurchenco, Peter
; TITLE OF INVENTION: Laminin 2 and Methods for Its Use
; FILE REFERENCE: 99-274-B
; CURRENT APPLICATION NUMBER: US/09/562,702A
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/155,945
; PRIOR FILING DATE: 1999-09-24
; PRIOR APPLICATION NUMBER: 60/143,289
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/139,198
; PRIOR FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: 60/131,720
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 5689
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (178)..(5535)
; NAME/KEY: sig_peptide
; LOCATION: (178)..(240)
JS-09-562-702A-17
Alignment Scores:
Pred. No.: 127
Score: 42.00
Length: 5689
Percent Similarity: 71.43%
Matches: 7
Best Local Similarity: 50.00%
Conservative: 3
Query Match: 53.16%
Mismatches: 4
Indels: 0
DB: 4
Gaps: 0
US-09-737-297-3 (1-16) x US-09-562-702A-17 (1-5689)
QY      3  GlySerThr***AspValTyxGlnAsnIleGlnTyxAlaGly 16
      |||||
Db      4083  GGCAGACAGCTGGAGTTTATCAAAAACCTCCGATATTCAGGG 4124
RESULT 4
US-09-561-818A-17
; Sequence 17, Application US/09561818A
; Patent No. 6638907
; GENERAL INFORMATION:
; APPLICANT: Korteasaa, Jariko
; APPLICANT: Tryggvason, Karl
; TITLE OF INVENTION: Laminin 8 and Methods For Its Use
; FILE REFERENCE: 99,274-D
; CURRENT APPLICATION NUMBER: US/09/561,818A
; CURRENT FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 5689
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (178)..(5535)
; NAME/KEY: sig_peptide
; LOCATION: (178)..(240)
US-09-561-818A-17
Alignment Scores:
Pred. No.: 127
Score: 42.00
Length: 5689
Percent Similarity: 71.43%
Matches: 7
Best Local Similarity: 50.00%
Conservative: 3
Query Match: 53.16%
Mismatches: 4
Indels: 0
DB: 4
Gaps: 0

```

```
Score:          42.00      Matches:          7
Percent Similarity:    71.43%   Conservative:        3
Best Local Similarity:  50.00%   Mismatches:         4
Query Match:           53.16%   Indels:             0
DB:                    4       Gaps:                 0

US-09-737-297-3 (1-16) X US-09-561-818A-17 (1-5689)

RESULT 5
US-08-961-527-45/c
; Sequence 45, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleo
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: P8340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11384 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-961-527-45
Alignment Scores:          293            Length:          11384
Pred. No.:                42.00          Matches:          7
Score:                     76.92%        Conservative:        3
Percent Similarity:        53.85%        Mismatches:         4
Best Local Similarity:     53.16%        Indels:             0
Query Match:               4              Gaps:                 0
DB:                                     0

US-09-737-297-3 (1-16) X US-08-961-527-45 (1-11384)

QY      2 GluglyseThr***AspValtyrGlnAsnIleGlnTyrAlagly 14
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      8894 GAGGGACTAGTCACGCCTATTATCCAAATAATCGACTAT 8856

RESULT 6
US-07-744-570B-1
; Sequence 1, Application US/07744570B
; Patent No. 5202249
; GENERAL INFORMATION:
```

APPLICANT: Kluepfel, D.
APPLICANT: Morosoli, R.
APPLICANT: Shareck, F.
TITLE OF INVENTION: Xylanase for Biobleaching
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Michael J. Bradley
STREET: 1200 South 47th Street
CITY: Box Number 4023
STATE: Richmond
COUNTRY: California
ZIP: 94804-0023
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44Mb storage
COMPUTER: IBM
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/744,570B
FILING DATE: 19910813
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 675 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: Single strand
TOPOLOGY: Circular
US-07-744-570B-1

Alignment Scores:
Pred. No.: 15.3 Length: 675
Score: 41.00 Matches: 6
Percent Similarity: 71.43% Conservative: 4
Best Local Similarity: 42.86% Mismatches: 4
Query Match: 51.90% Indels: 0
DB: 1 Gaps: 0

US-09-737-297-3 (1-16) x US-07-744-570B-1 (1-675)

QY 1 AlaGluGlySerThr***AspValTyrglnAsnIleGlnTy 14
Db 346 AGCGACGAGGACCTACGACATCTACGACGACCGGTAC 387

RESULT 7
US-09-453-702B-206
Sequence 206 Application US/09453702B
Patent No. 6365723
GENERAL INFORMATION:
APPLICANT: Blattner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 206:
SEQUENCE CHARACTERISTICS:
LENGTH: 43360
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 206:
US-09-453-702B-206

Alignment Scores:
Pred. No.: 2.38e-03 Length: 43360
Score: 41.00 Matches: 7
Percent Similarity: 62.50% Conservative: 3
Best Local Similarity: 43.75% Mismatches: 6
Query Match: 51.90% Indels: 0
DB: 4 Gaps: 0

US-09-737-297-3 (1-16) x US-09-453-702B-206 (1-43360)

QY 1 AlaGluGlySerThr***AspValTyrglnAsnIleGlnTyAlaGly 16
Db 30219 TGTCAGGTTCCACAGATACATTATCAGACTCTGCAATATCTCTGGC 30266

RESULT 8
US-09-453-702B-261
Sequence 261 Application US/09453702B
Patent No. 6365723
GENERAL INFORMATION:
APPLICANT: Blattner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 261:
SEQUENCE CHARACTERISTICS:

LENGTH: 45325 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (Genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 261:
US-09-453-702B-261

Alignment Scores:
Pred. No.: 2,51e+03 Length: 45325
Score: 41.00 Matches: 7
Percent Similarity: 62.50% Conservative: 3
Best Local Similarity: 43.75% Mismatches: 6
Query Match: 51.90% Indels: 0
DB: 4 Gaps: 0

US-09-737-297-3 (1-16) x US-09-453-702B-261 (1-45325)

2y 1 AlaGlucySerThr***ApValTyGlnAsnIleGlnTyRAlaGly 16
Db 31121 TCGTCAGGTTCACACAGATACATTATCAGACTCTGCATATCTGTGC 31168

RESULT 9

US-09-557-884-1

Sequence 1, Application US/09557884

Patent No. 6506581

GENERAL INFORMATION:

APPLICANT: Fleischmann et al.

TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/557,884

FILING DATE: 25-Apr-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/476,102

FILING DATE: JUN-5-1995

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks

REGISTRATION NUMBER: 41,971

REFERENCE/DOCKET NUMBER: PB186P3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-309-8504

TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-557-884-1

Alignment Scores:

Pred. No.: 2,04e+05 Length: 1830121

Score: 41.00 Matches: 7

Percent Similarity: 71.43% Conservative: 3

Best Local Similarity: 50.00% Mismatches: 4

Query Match: 51.90% Indels: 0

US-09-557-884-1

DB: 4 Gaps: 0

US-09-737-297-3 (1-16) x US-09-557-884-1 (1-1830121)

Qy 1 AlaGlucySerThr***ApValTyGlnAsnIleGlnTyR 14
Db 1728229 GCAAACGGGTCTATGGCACACGTTTTTCAGAACGTCGCTAT 1728270

RESULT 10

US-09-643-990A-1

Sequence 1, Application US/09643990A

Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann

Mark D. Adams

Owen White

Hamilton O. Smith

J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of

the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville,

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/643,990A

FILING DATE: 23-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,429

FILING DATE: 1995-06-07

APPLICATION NUMBER: 08/426,787

FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB186P1C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-610-5790

TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-643-990A-1

Alignment Scores:

Pred. No.: 2,04e+05 Length: 1830121

Score: 41.00 Matches: 7

Percent Similarity: 71.43% Conservative: 3

Best Local Similarity: 50.00% Mismatches: 4

Query Match: 51.90% Indels: 0

DB: 4 Gaps: 0

US-09-737-297-3 (1-16) x US-09-643-990A-1 (1-1830121)

Qy 1 AlaGlucySerThr***ApValTyGlnAsnIleGlnTyR 14
Db 1728229 GCAAACGGGTCTATGGCACACGTTTTTCAGAACGTCGCTAT 1728270

```

Db 806 GCCGAAGGCGTGAGTTGGACTACGTACCGTACCGTCCCTGCAGTACGTGGC 759

RESULT 13
US-08-961-527-235
; Sequence 235, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; FILING APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 235:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1766 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-235

Alignment Scores:
Pred. No.: 99.8 Length: 1766
Score: 39.50 Matches: 9
Percent Similarity: 55.00% Conservative: 2
Best Local Similarity: 45.00% Mismatches: 4
Query Match: 50.00% Indels: 5
DB: 4 Gaps: 1

US-09-737-297-3 (1-16) x US-08-961-527-235 (1-1766)
QY 2 GluglySerThr***AspValTyrGlnAsn-----lleGlnTyrAlaGly 16
DB 190 GAAGGAGAAAATATCGATTTGTATCAAAATAAAGGTCGCTTTTTTATACAATAATGCTGGT 249

RESULT 14
US-09-551-826D-13
; Sequence 13, Application US/09551826D
; Patent No. 6558939
; GENERAL INFORMATION:
; APPLICANT: No. 6558939regaard-Madsen, Mads
; APPLICANT: Ostergaard, Peter Rahbek
; APPLICANT: Christensen, Claus Bo Voge
; APPLICANT: Lassen, Soren Flensted
; TITLE OF INVENTION: No. 6558939el Ptoeases And Variants Thereof
; FILE REFERENCE: 5665-200-US
; CURRENT APPLICATION NUMBER: US/09/551,826D
; CURRENT FILING DATE: 2000-04-17
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.2

Alignment Scores:
Pred. No.: 81.3 Length: 1812
Score: 40.00 Matches: 7
Percent Similarity: 68.75% Conservative: 4
Best Local Similarity: 43.75% Mismatches: 5
Query Match: 50.63% Indels: 0
DB: 4 Gaps: 0

US-09-737-297-3 (1-16) x US-09-732-615-28 (1-1812)
QY 1 AlaGluGlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16
DB 398 ACTGAGATATTATATAGATCACTTCATATGCT 366

RESULT 12
US-09-732-615-28/c
; Sequence 28, Application US/09732615
; Patent No. 6632650
; GENERAL INFORMATION:
; APPLICANT: Chen, Mario W.
; APPLICANT: Chen, Qiong
; APPLICANT: Gibson, Katherine J.
; APPLICANT: Kostichka, Kristy N.
; APPLICANT: Thomas, Stuart M.
; APPLICANT: Nagarajan, Vasantha
; TITLE OF INVENTION: Genes Involved in Cyclododecanone Degradation Pathway
; FILE REFERENCE: BC1023 US NA
; CURRENT APPLICATION NUMBER: US/09/732,615
; CURRENT FILING DATE: 2000-12-08
; PRIOR FILING DATE: December 10, 1999
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Microsoft Office 97
; LENGTH: 1812
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Cdda Homolog
US-09-732-615-28

Alignment Scores:
Pred. No.: 14.9 Length: 447
Score: 40.00 Matches: 6
Percent Similarity: 90.91% Conservative: 4
Best Local Similarity: 54.55% Mismatches: 1
Query Match: 50.63% Indels: 0
DB: 4 Gaps: 0

US-09-737-297-3 (1-16) x US-09-621-976-13424 (1-447)
QY 5 Thr***AspValTyrGlnAsnIleGlnTyrAla 15
DB 398 ACTGAGATATTATATAGATCACTTCATATGCT 366

RESULT 11
US-09-621-976-13424/c
; Sequence 13424, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 13424
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-621-976-13424

Alignment Scores:
Pred. No.: 14.9 Length: 447
Score: 40.00 Matches: 6
Percent Similarity: 90.91% Conservative: 4
Best Local Similarity: 54.55% Mismatches: 1
Query Match: 50.63% Indels: 0
DB: 4 Gaps: 0

```

```

Db 806 GCCGAAGGCGTGAGTTGGACTACGTACCGTACCGTCCCTGCAGTACGTGGC 759

RESULT 13
US-08-961-527-235
; Sequence 235, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; FILING APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 235:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1766 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-235

Alignment Scores:
Pred. No.: 99.8 Length: 1766
Score: 39.50 Matches: 9
Percent Similarity: 55.00% Conservative: 2
Best Local Similarity: 45.00% Mismatches: 4
Query Match: 50.00% Indels: 5
DB: 4 Gaps: 1

US-09-737-297-3 (1-16) x US-08-961-527-235 (1-1766)
QY 2 GluglySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16
DB 190 GAAGGAGAAAATATCGATTTGTATCAAAATAAAGGTCGTTTTTATACAATAATGCTGGT 249

RESULT 14
US-09-551-826D-13
; Sequence 13, Application US/09551826D
; Patent No. 6558939
; GENERAL INFORMATION:
; APPLICANT: No. 6558939regaard-Madsen, Mads
; APPLICANT: Ostergaard, Peter Rahbek
; APPLICANT: Christensen, Claus Bo Voge
; APPLICANT: Lassen, Soren Flensted
; TITLE OF INVENTION: No. 6558939el Ptoeases And Variants Thereof
; FILE REFERENCE: 5665-200-US
; CURRENT APPLICATION NUMBER: US/09/551,826D
; CURRENT FILING DATE: 2000-04-17
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.2

Alignment Scores:
Pred. No.: 81.3 Length: 1812
Score: 40.00 Matches: 7
Percent Similarity: 68.75% Conservative: 4
Best Local Similarity: 43.75% Mismatches: 5
Query Match: 50.63% Indels: 0
DB: 4 Gaps: 0

US-09-737-297-3 (1-16) x US-09-732-615-28 (1-1812)
QY 1 AlaGluGlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16
DB 190 GAAGGAGAAAATATCGATTTGTATCAAAATAAAGGTCGTTTTTATACAATAATGCTGGT 249

```

```
; SEQ ID NO 13
; LENGTH: 939
; TYPE: DNA
; ORGANISM: Bacillus subtilis IS75
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(939)
; FEATURE:
; NAME/KEY: pro-peptide
; LOCATION: (103)..(279)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (280)..()
US-09-551-826D-13
Alignment Scores:
Pred. No.: 58.8 Length: 939
Score: 39.00 Matches: 7
Percent Similarity: 80.00% Conservative: 1
Best Local Similarity: 70.00% Mismatches: 2
Query Match: 49.37% Indels: 0
DB: 4 Gaps: 0

US-09-737-297-3 (1-16) x US-09-551-826D-13 (1-939)
2y 5 Thr***AspValTyrGlnAsnIleGIntYr 14
Db 898 ACGAAGCATGTATTCAACAATATTCATAT 927

RESULT 15
US-07-921-796-5/c
; Sequence 5, Application US/07921796
; Patent No. 5487990
; GENERAL INFORMATION:
; APPLICANT: Smith, John A.
; APPLICANT: Lee, Fang-Jen S.
; APPLICANT: Lin, Lee-Wen
; TITLE OF INVENTION: The Glucose-Regulated Promoter of Yeast
; TITLE OF INVENTION: Acetyl-CoA Hydrolase
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/921.796
; FILING DATE: 30-JUL-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanzo, Michael A.
; REGISTRATION NUMBER: 36,912
; REFERENCE/DOCKET NUMBER: 0609.1600003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2466 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; REGISTRATION NUMBER: 36,912
; REFERENCE/DOCKET NUMBER: 0609.1600003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2466 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 614..2191

US-09-737-297-3 (1-16) x US-09-551-826D-13 (1-939)
2y 5 Thr***AspValTyrGlnAsnIleGIntYr 14
Db 898 ACGAAGCATGTATTCAACAATATTCATAT 927

RESULT 16
US-07-921-796-7/c
; Sequence 7, Application US/07921796
; Patent No. 5487990
; GENERAL INFORMATION:
; APPLICANT: Smith, John A.
; APPLICANT: Lee, Fang-Jen S.
; APPLICANT: Lin, Lee-Wen
; TITLE OF INVENTION: The Glucose-Regulated Promoter of Yeast
; TITLE OF INVENTION: Acetyl-CoA Hydrolase
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/921.796
; FILING DATE: 30-JUL-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanzo, Michael A.
; REGISTRATION NUMBER: 36,912
; REFERENCE/DOCKET NUMBER: 0609.1600003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3089 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1237..2814
US-07-921-796-7
Alignment Scores:
Pred. No.: 249 Length: 3089
Score: 39.00 Matches: 7
Percent Similarity: 69.23% Conservative: 2
Best Local Similarity: 53.85% Mismatches: 4
Query Match: 49.37% Indels: 0
DB: 1 Gaps: 0

US-09-737-297-3 (1-16) x US-07-921-796-7 (1-3089)
Qy 2 GluGlySerThr***AspValTyrGlnAsnIleGIntYr 14
Db 980 GAGGGAACATGCACAGATGTTTATCAAGAACTCAATTC 942
```



```

RESULT 22
US-09-328-352-2901
; Sequence 2901, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3818
; LENGTH: 1632
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-3818

```



```

DB      1666 ACATTCTACTATATTTGATATACATATACATACAGGGG 1631
      RESULT 26
      US-08-449-609-6/c
      ; Sequence 6, Application US/08449609
      ; Patent No. 5952212
      ;
      GENERAL INFORMATION:
      APPLICANT: Moller, Niels P.H.
      APPLICANT: Moller, Karin B.
      APPLICANT: Moller, Axel
      APPLICANT: Ullrich, Axel
      TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
      PHOSPHATASE
      NUMBER OF SEQUENCES: 45
      CORRESPONDENCE ADDRESS:
      ADDRESSEE: PENNIE & EDMONDS
      STREET: 1155 Avenue of the Americas
      CITY: New York
      STATE: New York
      COUNTRY: U.S.A.
      ZIP: 10036-2711
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patent In Release #1.0, Version #1.25
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/449,609
      FILING DATE: 24-MAY-1995
      CLASSIFICATION: 435
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/036,210
      FILING DATE: 23-MAR-1993
      ATTORNEY/AGENT INFORMATION:
      NAME: Misrock, S. Leslie
      REGISTRATION NUMBER: 18,782
      REFERENCE/DOCKET NUMBER: 7683-025
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: 212-790-9090
      TELEFAX: 212-869-8864/9741
      TELEX: 66141 PENNIE
      INFORMATION FOR SEQ ID NO: 6:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 2173 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: unknown
      MOLECULE TYPE: cDNA
      FEATURE:
      NAME/KEY: misc feature
      LOCATION: 1630
      OTHER INFORMATION: /note= "N=x=unknown nucleotide"
      US-08-449-609-6
      Alignment Scores:
      Fred. No.: 260
      Score: 38.00
      Percent Similarity: 56.67%
      Best Local Similarity: 58.33%
      Query Match: 48.10%
      DB: 2
      US-09-737-297-3 (1-16) x US-08-449-609-6 (1-2173)
      Length: 2173
      Matches: 7
      Conservative: 1
      Mismatches: 4
      Indels: 0
      Gaps: 0
      QY 5 Thr***AspValtyrGlnAsnIleGlnTyrAlaGly 16
      ||| :||| ||||| ||||| |||||
      DB 1666 ACATTCTACTATATTTGATATACATATACAGGGG 1631
      RESULT 27
      US-09-361-096A-6/c
      ; Sequence 6, Application US/09361096A
      ; Patent No. 6492495
      ; GENERAL INFORMATION:
  
```

```

; TELEFAX: 212-869-8864/9741
;
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2309 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1766
; OTHER INFORMATION: /note= "N=x-unknown nucleotide"
; US-08-036-210-10
;
Alignment Scores:
Pred. No.: 280 Length: 2309
Score: 38.00 Matches: 7
Percent Similarity: 66.67% Conservative: 1
Best Local Similarity: 58.33% Mismatches: 4
Query Match: 48.10% Indels: 0
DB: 1 Gaps: 0
;
US-09-737-297-3 (1-16) x US-08-036-210-10 (1-2309)
;
QY 5 Thr***AspValtyrGlnAsnileglnTyralagly 16
;
DB 1802 ACATTACTATATATTGAATATACAATCAGGGGG 1767
;
RESULT 29
US-08-449-609-10/c
; Sequence 10, Application US/08449609
; Patent No. 5952212
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,609
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/036,210
; FILING DATE: 23-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Miarock, S. Leslie
; REGISTRATION NUMBER: 18,972
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2309 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
;

```

```

; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1756
; OTHER INFORMATION: /note= "N=x-unknown nucleotide"
US-08-449-609-10

Alignment Scores:
Pred. No.: 280
Score: 38.00
Length: 2309
Percent Similarity: 66.67% Matches: 7
Best Local Similarity: 58.33% Conservative: 1
Query Match: 48.10% Mismatches: 4
DB: 2 Indels: 0
Gaps: 0

US-09-737-297-3 (1-16) x US-08-449-609-10 (1-2309)
Qy 5 Thr***AspValTyGlnAsnIleGlnTyzAlaGly 16
|||::|||
Db 1802 ACATTCTACTATATTTGATATATACATACACAGGGG 1767

RESULT 30
US-09-361-096A-10/c
; Sequence 10, Application US/09361096A
; Patent No. 6492495
; GENERAL INFORMATION:
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; TITLE OF INVENTION: PTP-531: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/09/361,096A
; PRIOR FILING DATE: 1993-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 2309
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PTP-531D
; NAME/KEY: modified base
; LOCATION: (1)-(2309)
; OTHER INFORMATION: n = unknown nucleotide
; NAME/KEY: CDS
; LOCATION: (65)..(1030)
US-09-361-096A-10

Alignment Scores:
Pred. No.: 280
Score: 38.00
Length: 2309
Percent Similarity: 66.67% Matches: 7
Best Local Similarity: 58.33% Conservative: 1
Query Match: 48.10% Mismatches: 4
DB: 2 Indels: 0
Gaps: 0

US-09-737-297-3 (1-16) x US-09-361-096A-10 (1-2309)
Qy 5 Thr***AspValTyGlnAsnIleGlnTyzAlaGly 16
|||::|||
Db 1802 ACATTCTACTATATTTGATATATACATACACAGGGG 1767

RESULT 31
US-08-036-210-14/c
; Sequence 14, Application US/08036210
; Patent No. 5585233
; GENERAL INFORMATION:
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.

```

```

; NAME/KEY: CDS
; LOCATION: (92...139, 259...1414)
US-09-361-096A-14

Alignment Scores:
Pred. No.: 338 Length: 2692
Score: 38.00 Matches: 7
Percent Similarity: 66.67% Conservative: 1
Best Local Similarity: 58.33% Mismatches: 4
Query Match: 48.10% Indels: 0
DB: 4 Gaps: 0

US-09-737-297-3 (1-16) x US-09-361-096A-14 (1-2692)

QY 5 Thr**AspValtyrGlnAsnleGlnTyAlaGly 16
Db 2185 ACATTTACTATATATTGAATATACATACAGGGG 2150

RESULT 34
US-09-107-532A-681
; Sequence 681, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 681:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2784 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...2784
; SEQUENCE DESCRIPTION: SEQ ID NO: 681:
US-09-107-532A-681

Alignment Scores:
Pred. No.: 352 Length: 2784

```

Score: 38.00 Matches: 7
Percent Similarity: 80.00% Conservative: 1
Best Local Similarity: 70.00% Mismatches: 2
Query Match: 48.10% Indels: 0
DB: 4 Gaps: 0

US-09-737-297-3 (1-16) x US-09-107-532A-681 (1-2784)

QY 4 SerThr***AspValTyrGlnAsnIleGln 13
DB 2041 AGCTTAGAGATATTTATCAAAATATCCAG 2070

RESULT 35

US-08-936-165A-194
; Sequence 194, Application US/08936165A
; Patent No. 6348582
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Burnham, Martin
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Lonetto, Michael
; APPLICANT: Nicholas, Richard
; APPLICANT: Pratt, Julie
; APPLICANT: Reichard, Richard
; APPLICANT: Rosenberg, Martin
; APPLICANT: Ward, Judith
; TITLE OF INVENTION: Polypeptides and Their Uses
; NUMBER OF SEQUENCES: 534
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,165A
; FILING DATE: 24-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/027,032
; FILING DATE: 24-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50549
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 194:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3191 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-08-936-165A-194

Alignment Scores:
Pred. No.: 415 Length: 3191
Score: 38.00 Matches: 6
Percent Similarity: 73.33% Conservative: 5
Best Local Similarity: 40.00% Mismatches: 4
Query Match: 48.10% Indels: 0
DB: 4 Gaps: 0

US-09-737-297-3 (1-16) x US-08-936-165A-194 (1-3191)
QY 2 GluGlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16
DB 1873 GATGGCAGCACTATCGATTGTGATGAAGTATTAAAGAAACAGGT 1917

RESULT 36

US-08-036-210-21/c
; Sequence 21, Application US/08036210
; Patent No. 5595233
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/036,210
; FILING DATE: 23-MAR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7883-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 3430-
; OTHER INFORMATION: /note="N=x-unknown nucleotide"
; US-08-036-210-21

Alignment Scores:
Pred. No.: 541 Length: 3973
Score: 38.00 Matches: 7
Percent Similarity: 66.67% Conservative: 1
Best Local Similarity: 58.33% Mismatches: 4
Query Match: 48.10% Indels: 0
DB: 1 Gaps: 0

US-09-737-297-3 (1-16) x US-08-036-210-21 (1-3973)

QY 5 Thr***AspValTyrGlnAsnIleGlnTyrAlaGly 16
DB 3466 ACATTTACTATATTTGATATACATACAGGGG 3431

RESULT 37

US-08-449-609-21/c
; Sequence 21, Application US/08449609
; Patent No. 5952212

GENERAL INFORMATION:
APPLICANT: Moller, Niels P.H.
APPLICANT: Moller, Karin B.
APPLICANT: Ullrich, Axel
TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
PHOSPHATASE
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/449,609
FILING DATE: 24-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 08/036,210
FILING DATE: 23-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7683-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 3973 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3430
OTHER INFORMATION: /note= "N=x=unknown nucleotide"
US-08-449-609-21

Alignment Scores:
Pred. No.: 541 Length: 3973
Score: 38.00 Matches: 7
Percent Similarity: 66.67% Conservative: 1
Best Local Similarity: 58.33% Mismatches: 4
Query Match: 48.10% Indels: 0
DB: 2 Gaps: 0

US-09-737-297-3 (1-16) x US-08-449-609-21 (1-3973)
QY 5 Thr***AspValtyrGlnAsnIleGlnTyrAlaGly 16
DB 3466 ACATTACTATATATTGATATACATACAGGGG 3431

RESULT 38
US-09-361-096A-21/C
Sequence 21, Application US/09361096A
Patent No. 6492495
GENERAL INFORMATION:
APPLICANT: MOLLER, NIELS P.H.
APPLICANT: MOLLER, KARIN B.
APPLICANT: ULLRICH, AXEL
TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
FILE REFERENCE: 038602/0886
CURRENT APPLICATION NUMBER: US/09/361,096A
CURRENT FILING DATE: 1999-07-26

PRIOR APPLICATION NUMBER: 08/449,609
PRIOR FILING DATE: 1995-05-24
PRIOR APPLICATION NUMBER: 08/036,210
PRIOR FILING DATE: 1995-03-23
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 21
LENGTH: 3973
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: PTPS31-RD#2
NAME/KEY: modified base
LOCATION: (1)..(3973)
OTHER INFORMATION: n = unknown nucleotide
NAME/KEY: CDS
LOCATION: (1)..(2894)
US-09-361-096A-21

Alignment Scores:
Pred. No.: 541 Length: 3973
Score: 38.00 Matches: 7
Percent Similarity: 66.67% Conservative: 1
Best Local Similarity: 58.33% Mismatches: 4
Query Match: 48.10% Indels: 0
DB: 4 Gaps: 0

US-09-737-297-3 (1-16) x US-09-361-096A-21 (1-3973)
QY 5 Thr***AspValtyrGlnAsnIleGlnTyrAlaGly 16
DB 3466 ACATTACTATATATTGATATACATACAGGGG 3431

RESULT 39
US-08-956-171E-234
Sequence 234, Application US/08956171E
Patent No. 6593114
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
Gil H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 05/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224

Alignment Scores:

FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 1.16e+06 4403765
Score: 38.00 7
Percent Similarity: 90.00% 2
Best Local Similarity: 70.00% 1
Query Match: 48.10% 0
DB: 3 0

US-09-737-297-3 (1-16) x US-09-103-840A-2 (1-4403765)

Qy 7 AspValTyrGlnAsnIleGlnTyAlaGly 16

Db 4320417 GATGTGCACCGATGATCCAGCAGCGGGC 4320388

RESULT 43

US-09-103-840A-1

; Sequence 1, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 4411529

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; OTHER INFORMATION: H37RV

US-09-103-840A-1

Alignment Scores:

Pred. No.: 1.16e+06 4411529
Score: 38.00 9
Percent Similarity: 62.50% 1
Best Local Similarity: 56.25% 6
Query Match: 48.10% 0
DB: 3 0

US-09-737-297-3 (1-16) x US-09-103-840A-1 (1-4411529)

Qy 1 AlaGluGlySerThr***AspValTyrGlnAsnIleGlnTyAlaGly 16

Db 967082 GCCGAGCGAGTTACCGCGATGTATCAGCTGCCAGTTCGTCGGG 967129

RESULT 44

US-09-103-840A-1/c

; Sequence 1, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Alignment Scores:
Pred. No.: 1.16e+06 4411529
Score: 38.00 7
Percent Similarity: 90.00% 2
Best Local Similarity: 70.00% 1
Query Match: 48.10% 0
DB: 3 0

US-09-737-297-3 (1-16) x US-09-103-840A-1 (1-4411529)

Qy 7 AspValTyrGlnAsnIleGlnTyAlaGly 16

Db 4328163 GATGTGCACCGATGATCCAGCAGCGGGC 4328134

RESULT 45

US-09-189-060B-46

; Sequence 46, Application US/09189060B

; Patent No. 6270968

; GENERAL INFORMATION:

; APPLICANT: Dalboge, Henrik

; APPLICANT: Sandal, Thomas

; APPLICANT: Kauppinen, Markus

; APPLICANT: Borge, Diderichsen

; TITLE OF INVENTION: Method Of Providing No. 6270968el DNA Sequences

; FILE REFERENCE: 4772.204-US

; CURRENT APPLICATION NUMBER: US/09/189,060B

; CURRENT FILING DATE: 1998-11-10

; PRIOR APPLICATION NUMBER: PCT/DK97/00216

; PRIOR FILING DATE: 1997-05-12

; NUMBER OF SEQ ID NOS: 74

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 46

; LENGTH: 234

; TYPE: DNA

; ORGANISM: Hybrid

US-09-189-060B-46

Alignment Scores:

Pred. No.: 28.1 234
Score: 37.00 5
Percent Similarity: 69.23% 4
Best Local Similarity: 38.46% 4
Query Match: 46.84% 0
DB: 3 0

US-09-737-297-3 (1-16) x US-09-189-060B-46 (1-234)

Qy 2 GluGlySerThr***AspValTyrGlnAsnIleGlnTy 14

Db 11 GACGGGGGACGATGAYATACGAGACCCCGCTAC 49

RESULT 46

US-09-189-060B-54

; Sequence 54, Application US/09189060B

; Patent No. 6270968

; GENERAL INFORMATION:

; APPLICANT: Dalboge, Henrik

; APPLICANT: Sandal, Thomas

; APPLICANT: Kauppinen, Markus

; APPLICANT: Borge, Diderichsen

; TITLE OF INVENTION: Method Of Providing No. 6270968el DNA Sequences

; FILE REFERENCE: 4772.204-US

; CURRENT APPLICATION NUMBER: US/09/189,060B

; CURRENT FILING DATE: 1998-11-10

; PRIOR APPLICATION NUMBER: PCT/DK97/00216

; PRIOR FILING DATE: 1997-05-12

NUMBER OF SEQ ID NOS: 74
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 54
LENGTH: 234
TYPE: DNA
ORGANISM: Hybrid
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(234)
OTHER INFORMATION: n = a, c, g or t
US-09-189-060B-54

Alignment Scores:
Pred. No.: 28.1 Length: 234
Score: 37.00 Matches: 5
Percent Similarity: 69.23% Conservative: 4
Best Local Similarity: 38.46% Mismatches: 4
Query Match: 46.84% Indels: 0
DB: 3 Gaps: 0

US-09-737-297-3 (1-16) x US-09-189-060B-54 (1-234)

Qy 2 GluGlySerThr***AspValTyGlnAsnIleGlnTyR 14
Db 11 GACGGGGGACGTACGACATATACGAGACTACTCGTTAC 49

RESULT 47
US-09-543-681A-2067
Sequence 2067, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 2067
LENGTH: 240
TYPE: DNA
ORGANISM: Proteus mirabilis
US-09-543-681A-2067

Alignment Scores:
Pred. No.: 28.9 Length: 240
Score: 37.00 Matches: 6
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 75.00% Mismatches: 0
Query Match: 46.84% Indels: 0
DB: 4 Gaps: 0

US-09-737-297-3 (1-16) x US-09-543-681A-2067 (1-240)

Qy 7 AspValTyGlnAsnIleGlnTyR 14
Db 208 GATGCTATGAAATATATAATAT 231

RESULT 48
US-09-711-164-4/c
Sequence 4, Application US/09711164
Patent No. 6589738
GENERAL INFORMATION:
APPLICANT: Forsyth, R. Alllyn
APPLICANT: Ohlsen, Judith
APPLICANT: Zyskind, Judith
TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
FILE REFERENCE: ELITRA.008A
CURRENT APPLICATION NUMBER: US/09/711,164
CURRENT FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: US 60/164415

PRIOR FILING DATE: 1999-11-9
NUMBER OF SEQ ID NOS: 469
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 382
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(382)
OTHER INFORMATION: n = A,T,C or G
US-09-711-164-4

Alignment Scores:
Pred. No.: 50.8 Length: 382
Score: 37.00 Matches: 5
Percent Similarity: 87.50% Conservative: 2
Best Local Similarity: 62.50% Mismatches: 1
Query Match: 46.84% Indels: 0
DB: 4 Gaps: 0

US-09-737-297-3 (1-16) x US-09-711-164-4 (1-382)

Qy 7 AspValTyGlnAsnIleGlnTyR 14
Db 68 GATATTATCAGATCTGTATTAC 45

RESULT 49
US-09-621-976-18487/c
Sequence 18487, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S. J.Y.
TITLE OF INVENTION: ESTS and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 18487
LENGTH: 493
TYPE: DNA
ORGANISM: Homo sapiens
US-09-621-976-18487

Alignment Scores:
Pred. No.: 59.2 Length: 493
Score: 37.00 Matches: 7
Percent Similarity: 72.73% Conservative: 1
Best Local Similarity: 63.64% Mismatches: 3
Query Match: 46.84% Indels: 0
DB: 4 Gaps: 0

US-09-737-297-3 (1-16) x US-09-621-976-18487 (1-493)

Qy 2 GluGlySerThr***AspValTyGlnAsnIle 12
Db 363 GAAAGGAAACACCTGATATATATACCAACATT 331

RESULT 50
US-08-275-526C-26
Sequence 26, Application US/08275526C
Patent No. 6180382
GENERAL INFORMATION:
APPLICANT: DE BUYL, ERIC
APPLICANT: LAHAYE, ANDR E
APPLICANT: LEDOUX, PIERRE
APPLICANT: AMORY, ANTOINE
APPLICANT: DETROZ, REN
APPLICANT: ANDRE, CHRISTOPHE
APPLICANT: VETTER, ROMAN

```

; TITLE OF INVENTION: XYLANASE DERIVED FROM A BACILLUS SPECIES,
; TITLE OF INVENTION: EXPRESSION VECTORS FOR SUCH XYLANASE AND
; TITLE OF INVENTION: OTHER PROTEINS, HOST ORGANISMS THEREFOR AND
; TITLE OF INVENTION: USE THEREOF
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WILLIAM BRINKS HOFER GILSON & LIONE, P.C.
; STREET: 2000 K St., N.W., Suite 200
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/275,526C
; FILING DATE: 15-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gadiano, Wilhem F.
; REGISTRATION NUMBER: 37,136
; REFERENCE/DOCKET NUMBER: 4121-49
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 429-0625
; TELEFAX: (202) 293-0625
; TELEX: 650 383 5605
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-08-275-526C-26

Alignment Scores:
Pred. No.: 87.8 Length: 600
Score: 37.00 Matches: 5
Percent Similarity: 76.92% Conservative: 3
Best Local Similarity: 38.46% Mismatches: 0
Query Match: 46.84% Indels: 0
DB: 3 Gaps: 0

US-09-737-297-3 (1-16) x US-08-275-526C-26 (1-600)
Qy 1 AlaGluGlySerThr***AspValTyrGlnAsnIleGln 13
Db 343 GCCGATGGAGGCACATATGACATATATGAAAGCGTCCGT 381

Search completed: March 1, 2004, 12:32:41
Job time : 3547 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 1, 2004, 10:17:00 ; Search time 264 Seconds
(without alignments)

218.634 Million cell updates/sec

Title: US-09-737-297-3

Perfect score: 79

Sequence: 1 ARGSTXDYQNIQVAG 16

Scoring table:

BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Dgapop 6.0, Dgapext 7.0

Searched: 2353733 seqs, 180373377 residues

Total number of hits satisfying chosen parameters: 4707466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Command line parameters:

-MODEL=frame+*_p2n.model -DEV=xlp
-O=/cgn2_1/USPFC_pool_p/US09737297/runat_01032004_085324_3520/app_query.fasta_1.199
-DB=Published Applications NA -OFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=150 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=50 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09737297 @CNC 1 1 333 @runat 01032004_085324_3520
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FCGPOP=6 -FCGPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:**

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55.7	346	15	US-10-242-535A-42365	Sequence 42365, A
2	55.7	1358	15	US-10-027-632-124226	Sequence 124226, A
3	55.7	684973	9	US-09-263-959-1	Sequence 1, Appli
4	54.4	493	12	US-10-424-599-103561	Sequence 103561, A
5	54.4	1143	15	US-10-359-493-42043	Sequence 42043, A
6	53.2	539	12	US-10-424-599-130893	Sequence 130893, A
7	53.2	724	12	US-10-282-122A-37042	Sequence 37042, A
8	53.2	5329	14	US-10-037-182-11	Sequence 11, Appli
9	53.2	5689	14	US-10-037-182-9	Sequence 9, Appli
10	53.2	11384	12	US-10-158-844-45	Sequence 45, Appli
11	51.9	440	15	US-10-027-632-183275	Sequence 183275, A
12	51.9	445	11	US-09-864-408A-2561	Sequence 2561, Ap
13	51.9	1230	9	US-09-815-242-7173	Sequence 7173, Ap
14	51.9	2010	14	US-10-289-757-17	Sequence 17, Appli
15	51.9	43360	14	US-10-114-170-206	Sequence 206, App
16	51.9	45325	14	US-10-114-170-261	Sequence 261, App
17	51.9	1830121	14	US-10-329-960-1	Sequence 1, Appli
18	51.9	1830121	15	US-10-329-670-1	Sequence 1, Appli
19	50.6	415	9	US-09-960-352-11946	Sequence 11946, A
20	50.6	1371	9	US-09-815-242-6773	Sequence 6773, Ap
21	50.6	1554	10	US-09-759-1308-182	Sequence 182, App
22	50.6	1773	14	US-10-276-629-7	Sequence 7, Appli
23	50.6	1776	12	US-10-282-122A-29515	Sequence 29515, A
24	50.6	1793	9	US-09-925-300-644	Sequence 644, App
25	50.6	1812	14	US-10-273-051-28	Sequence 28, Appli
26	50.6	1812	15	US-10-129-518-28	Sequence 28, Appli
27	50.6	2151	9	US-09-845-157-1	Sequence 1, Appli
28	50.6	2467	10	US-09-759-1308-181	Sequence 181, App
29	50.6	3691	13	US-10-044-090-796	Sequence 796, App
30	50.6	2140405	15	US-10-027-632-76212	Sequence 76212, A
31	50.0	1766	12	US-10-158-844-235	Sequence 235, App
32	49.4	50	15	US-10-131-827-6245	Sequence 6245, Ap
33	49.4	563	15	US-10-027-632-87366	Sequence 87366, A
34	49.4	847	15	US-10-027-632-138455	Sequence 138455, A
35	49.4	847	15	US-10-027-632-157309	Sequence 157309, A
36	49.4	1320	15	US-10-369-493-32618	Sequence 32618, A
37	49.4	1872	12	US-10-282-122A-24198	Sequence 24198, A
38	49.4	1872	15	US-10-398-221-847	Sequence 847, App
39	49.4	1872	15	US-10-398-221-2762	Sequence 2762, Ap
40	49.4	2000	9	US-09-887-576-118	Sequence 118, App
41	49.4	2000	9	US-09-938-842A-3048	Sequence 3048, Ap
42	49.4	2000	9	US-09-938-842A-3386	Sequence 3386, Ap
43	49.4	2000	11	US-09-938-842A-3048	Sequence 3048, Ap
44	49.4	2000	11	US-09-938-842A-3386	Sequence 3386, Ap
45	49.4	2598	12	US-10-282-122A-14296	Sequence 14296, A
46	49.4	3378	14	US-10-102-806-304	Sequence 304, App
47	49.4	3907	15	US-10-264-049-205	Sequence 205, App
48	49.4	23378	9	US-09-764-847-1514	Sequence 1514, Ap
49	49.4	23378	14	US-10-092-154-1514	Sequence 1514, Ap
50	49.4	326014	9	US-09-731-231A-3	Sequence 3, Appli
51	49.4	1691139	14	US-10-067-514-1	Sequence 1, Appli
52	49.4	1691139	15	US-10-419-723-1	Sequence 1, Appli
53	49.4	450	15	US-10-242-535A-30666	Sequence 30666, A
54	48.7	2748	9	US-09-764-898-53	Sequence 53, Appli
55	48.7	20272	9	US-09-908-711-145	Sequence 145, App
56	48.7	20272	10	US-09-764-891-6215	Sequence 6215, Ap
57	48.7	25619	9	US-09-908-711-143	Sequence 143, App
58	48.7	25619	9	US-09-764-898-302	Sequence 302, App
59	48.7	25619	10	US-09-764-891-6213	Sequence 6213, Ap
60	48.7	42595	11	US-09-984-429-469	Sequence 469, App
61	48.1	318	9	US-09-864-761-17871	Sequence 17871, A
62	48.1	378	11	US-09-864-408A-3841	Sequence 3841, Ap
63	48.1	401	9	US-09-795-668-122	Sequence 122, App
64	48.1	401	9	US-09-946-807-122	Sequence 122, App
65	48.1	401	9	US-09-867-701-9384	Sequence 9384, Ap
66	48.1	445	9	US-09-864-761-1101	Sequence 1101, Ap
67	48.1	470	9	US-10-191-803-354	Sequence 354, App
68	48.1	536	15	US-10-027-632-247735	Sequence 247735, A
69	48.1	598	15	US-10-027-632-192213	Sequence 192213, A
70	48.1	620	15	US-09-070-927A-780	Sequence 780, App
71	48.1	677	9	US-10-027-632-21424	Sequence 21424, A
72	48.1	702	15	US-10-027-632-21424	Sequence 21424, A
73	48.1	941	12	US-10-424-599-65055	Sequence 65055, A

74 38 48.1 942 14 US-10-213-990-71 Sequence 71, Appl
75 38 48.1 972 12 US-10-282-122A-18078 Sequence 18078, A
76 38 48.1 1002 14 US-10-213-990-70 Sequence 70, Appl
77 38 48.1 1012 9 US-09-815-242-1919 Sequence 1919, Appl
78 38 48.1 1012 12 US-10-282-122A-4438 Sequence 4438, Ap
79 38 48.1 1013 9 US-09-815-242-1876 Sequence 1876, Ap
80 38 48.1 1013 9 US-09-815-242-1877 Sequence 1877, Ap
81 38 48.1 1013 12 US-10-282-122A-4435 Sequence 4435, Ap
82 38 48.1 1013 12 US-10-282-122A-4440 Sequence 4440, Ap
83 38 48.1 1283 12 US-10-425-114-21156 Sequence 21156, A
84 38 48.1 1335 12 US-10-282-122A-21122 Sequence 21122, A
85 38 48.1 1494 12 US-10-282-122A-7771 Sequence 7771, Ap
86 38 48.1 1494 12 US-10-358-917-13 Sequence 13, Appl
87 38 48.1 2109 14 US-10-314-232-6 Sequence 6, Appl
88 38 48.1 2109 14 US-10-314-232-10 Sequence 10, Appl
89 38 48.1 2109 14 US-10-314-232-14 Sequence 14, Appl
90 38 48.1 2109 14 US-10-314-232-14 Sequence 14, Appl
91 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
92 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
93 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
94 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
95 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
96 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
97 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
98 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
99 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
100 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
101 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
102 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
103 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
104 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
105 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
106 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
107 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
108 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
109 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
110 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
111 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
112 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
113 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
114 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
115 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
116 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
117 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
118 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
119 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
120 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
121 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
122 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
123 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
124 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
125 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
126 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
127 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
128 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
129 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
130 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
131 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
132 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
133 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
134 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
135 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
136 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
137 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
138 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
139 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
140 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
141 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
142 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
143 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
144 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
145 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl
146 38 48.1 3191 9 US-09-939-980-194 Sequence 194, Appl

ALIGNMENTS

RESULT 1
US-10-242-535A-42365
; Sequence 42365, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 42365
; LENGTH: 346
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-42365
Alignment Scores:
Pred. No.: 5.8 Length: 346
Score: 44.00 Matches: 7
Percent Similarity: 78.57% Conservatives: 4
Best Local Similarity: 50.00% Mismatches: 3
Query Match: 55.70% Indels: 0
DB: 15 Gaps: 0
US-09-737-297-3 (1-16) x US-10-242-535A-42365 (1-346)
QY 2 GluglySerThr***ApValTyGlnAsnIleGlnTyzAla 15
DB 82 GATGTTCCACAGTGAATTCACCAATTCAGTATAGT 123
RESULT 2
US-10-027-632-124226
; Sequence 124226, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002

Sequence 129, App
Sequence 5, Appli
Sequence 9, Appli
Sequence 9611, Ap

```

; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 124226
; LENGTH: 1358
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-124226

Alignment Scores:
Pred. No.: 31.2 Length: 1358
Score: 44.00 Matches: 10
Percent Similarity: 73.33% Conservative: 1
Best Local Similarity: 66.67% Mismatches: 4
Query Match: 55.70% Indels: 0
DB: 15 Gaps: 0

US-09-737-297-3 (1-16) x US-10-027-632-124226 (1-1358)

2y 2 GluclySerThr***AspValTyGlnAsnIleGlnTyraGly 16
Db 503 GAAGATCTACTCTCAATGTCATGTCATATCCAGTCAGCTGGG 547

RESULT 3
US-09-263-959-1/c
; Sequence 1, Application US/09263959
; Patent No. US20020150891A1
; GENERAL INFORMATION:
; APPLICANT: Hood, Leroy E.
; APPLICANT: Rowen, Lee
; APPLICANT: Koop, Ben F.
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH UTIL
; NUMBER OF SEQUENCES: 1279
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/263,959
; FILING DATE: 05-MAR-1999
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McMaisters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 920010.426C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 684973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-263-959-1

Alignment Scores:
Pred. No.: 6.52e+04 Length: 684973
Score: 44.00 Matches: 10
Percent Similarity: 73.33% Conservative: 1
Best Local Similarity: 66.67% Mismatches: 4
Query Match: 55.70% Indels: 0
DB: 9 Gaps: 0

US-09-737-297-3 (1-16) x US-09-263-959-1 (1-684973)

; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 124226
; LENGTH: 1358
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-124226

Alignment Scores:
Pred. No.: 31.2 Length: 1358
Score: 44.00 Matches: 10
Percent Similarity: 73.33% Conservative: 1
Best Local Similarity: 66.67% Mismatches: 4
Query Match: 55.70% Indels: 0
DB: 15 Gaps: 0

US-09-737-297-3 (1-16) x US-10-027-632-124226 (1-1358)

2y 2 GluclySerThr***AspValTyGlnAsnIleGlnTyraGly 16
Db 503 GAAGATCTACTCTCAATGTCATGTCATATCCAGTCAGCTGGG 547

RESULT 4
US-10-424-599-103561/c
; Sequence 103561, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 103561
; LENGTH: 493
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(493)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_64533C.1
US-10-424-599-103561

Alignment Scores:
Pred. No.: 14.4 Length: 493
Score: 43.00 Matches: 7
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 87.50% Mismatches: 0
Query Match: 54.43% Indels: 0
DB: 12 Gaps: 0

US-09-737-297-3 (1-16) x US-10-424-599-103561 (1-493)

QY 7 AspValTyGlnAsnIleGlnTyraGly 14
Db 411 GATGTATATCAAAAGCTACATAT 388

RESULT 5
US-10-369-493-42043
; Sequence 42043, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 42043
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: Lactococcus lactis
US-10-369-493-42043

Alignment Scores:
Pred. No.: 40.6 Length: 1143
Score: 43.00 Matches: 8
```

Percent Similarity: 76.92% Conservative: 2
Best Local Similarity: 61.54% Mismatches: 3
Query Match: 54.43% Indels: 0
DB: 15 Gaps: 0

US-09-737-297-3 (1-16) x US-10-369-493-42043 (1-1143)

QY 2 GluGlySerThr***AspValTyrGlnAsnIleGlnTyr 14
DB 298 GAAGGAACTCTCACGCTGTTTATCAAAATATCTTAT 336

RESULT 6

US-10-424-599-130893
; Sequence 130893, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David X
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 130893
; LENGTH: 539
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_89203C.1
US-10-424-599-130893

Alignment Scores:
Pred. No.: 26 Length: 539
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 4
Best Local Similarity: 60.00% Mismatches: 0
Query Match: 53.16% Indels: 0
DB: 12 Gaps: 0

US-09-737-297-3 (1-16) x US-10-424-599-130893 (1-539)

QY 7 AspValTyrGlnAsnIleGlnTyrAlaGly 16
DB 64 GATATCTATCAGACATCGATATCTGCT 93

RESULT 7

US-10-282-122A-37042
; Sequence 37042, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA-034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37042
; LENGTH: 724
; TYPE: DNA
; ORGANISM: Salmonella paratyphi A
US-10-282-122A-37042

Alignment Scores:
Pred. No.: 37.3 Length: 724
Score: 42.00 Matches: 7
Percent Similarity: 59.23% Conservative: 2
Best Local Similarity: 53.85% Mismatches: 4
Query Match: 53.16% Indels: 0
DB: 12 Gaps: 0

US-09-737-297-3 (1-16) x US-10-282-122A-37042 (1-724)

QY 2 GluGlySerThr***AspValTyrGlnAsnIleGlnTyr 14
DB 57 GATGGCGGCACAGAGATGTTTATCTGATTCGAAAT 95

RESULT 8

US-10-037-182-11
; Sequence 11, Application US/10037182
; Publication No. US20030044899A1
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Doi, Masayuki
; APPLICANT: Thyboll, Jill
; TITLE OF INVENTION: Recombinant Laminin 10
; FILE REFERENCE: 99-274-P
; CURRENT APPLICATION NUMBER: US/10/037,182
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,449
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/279,282
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 5329
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(5175)
US-10-037-182-11

Alignment Scores:

Pred. No.: 435 Length: 5329
Score: 42.00 Matches: 7
Percent Similarity: 71.43% Conservative: 3
Best Local Similarity: 50.00% Mismatches: 4
Query Match: 53.16% Indels: 0
DB: 14 Gaps: 0

CLASSIFICATION	SEQUENCE	SEQUENCE DESCRIPTION	SEQ ID NO	LENGTH	SCORE	PERCENT SIMILARITY	BEST LOCAL SIMILARITY	QUERY MATCH	DB
US-09-737-297-3 (1-16) x US-10-037-182-11 (1-5329)	3 GlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16								
Db	3723 GGCAGACAGCTGGAGTTTATCAAAAACCTCCGATATTCAGGG 3764								
RESULT 9	US-10-037-182-9	SEQUENCE 9, Application US/10037182							
	Publication No. US20030044899A1	GENERAL INFORMATION:							
	APPLICANT: Tryggvason, Karl	APPLICANT: Doi, Masayuki							
	APPLICANT: Thyboll, Jill	TITLE OF INVENTION: Recombinant Laminin 10							
	FILE REFERENCE: 99-274-F	CURRENT APPLICATION NUMBER: US/10/037,182							
	CURRENT FILING DATE: 2001-12-21	PRIOR APPLICATION NUMBER: 60/257,449							
	PRIOR FILING DATE: 2000-12-21	PRIOR APPLICATION NUMBER: 60/279,282							
	PRIOR FILING DATE: 2001-03-28	NUMBER OF SEQ ID NOS: 36							
	SOFTWARE: PatentIn Ver. 2.0	SEQ ID NO 9							
	LENGTH: 5689	TYPE: DNA							
	ORGANISM: Mus musculus	FEATURE:							
	NAME/KEY: CDS	LOCATION: (178)..(5535)							
	NAME/KEY: sig_peptide	LOCATION: (178)..(240)							
US-10-037-182-9									
Alignment Scores:									
Pred. No.:	471	Length:	5689						
Score:	42.00	Matches:	7						
Percent Similarity:	71.43%	Conservative:	3						
Best Local Similarity:	50.00%	Mismatches:	4						
Query Match:	53.16%	Indels:	0						
DB:	14	Gaps:	0						
US-09-737-297-3 (1-16) x US-10-037-182-9 (1-5689)	3 GlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16								
Db	4083 GGCAGACAGCTGGAGTTTATCAAAAACCTCCGATATTCAGGG 4124								
RESULT 10	US-10-158-844-45/c	SEQUENCE 45, Application US/10158844							
	Publication No. US20040029118A1	GENERAL INFORMATION:							
	APPLICANT: Kunsch et al.	TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences							
	NUMBER OF SEQUENCES: 391	CORRESPONDENCE ADDRESS:							
	ADDRESSEE: Human Genome Sciences, Inc.	STREET: 9410 Key West Avenue							
	CITY: Rockville	STATE: Maryland							
	COUNTRY: USA	ZIP: 20850							
	COMPUTER READABLE FORM:	MEDIUM TYPE: CD-R							
	COMPUTER: Dell Latitude Pentium 3	OPERATING SYSTEM: Windows 98							
	SOFTWARE: ASCII Text	CURRENT APPLICATION DATA:							
	APPLICATION NUMBER: US/10/158,844	FILING DATE: 03-Jun-2002							
US-09-737-297-3 (1-16) x US-10-158-844-45 (1-11384)	2 GluGlySerThr***AspValTyrGlnAsnIleGlnTyr 14								
Db	8894 GAGGGAGCTAGTCACGCTATTATACAAAATATCGACTAT 8856								
RESULT 11	US-10-027-632-183275	SEQUENCE 183275, Application US/10027632							
	Publication No. US20030204075A9	GENERAL INFORMATION:							
	APPLICANT: Wang, David G.	TITLE OF INVENTION: Identification and Mapping of Single Nucleotide							
	FILE REFERENCE: 108827.129	CURRENT APPLICATION NUMBER: US/10/027,632							
	CURRENT FILING DATE: 2002-04-30	PRIOR APPLICATION NUMBER: US 60/218,006							
	PRIOR FILING DATE: 2000-07-12	PRIOR APPLICATION NUMBER: US 60/198,676							
	PRIOR FILING DATE: 2000-04-20	PRIOR APPLICATION NUMBER: US 60/193,483							
	PRIOR FILING DATE: 2000-03-29	PRIOR APPLICATION NUMBER: US 60/185,218							
	PRIOR FILING DATE: 2000-02-24	PRIOR APPLICATION NUMBER: US 60/167,363							
	PRIOR FILING DATE: 1999-11-23	PRIOR APPLICATION NUMBER: US 60/156,358	</						

```
US-09-737-297-3 (1-16) x US-10-027-632-183275 (1-440)
QY 2 GluGlySerThr***AspValTyrglnAsnIleGlnTyrAla 15
DB 42 GAGGGTCAGTCATAGATGCCTATGAGATATATATGCGAGCT 83
RESULT 12
US-09-864-408A-2561
; Sequence 2561, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shinkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2561
; LENGTH: 445
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-864-408A-2561
Alignment Scores:
Pred. No.: 33.1 Length: 445
Score: 41.00 Matches: 7
Percent Similarity: 84.62% Conservative: 4
Best Local Similarity: 53.85% Mismatches: 2
Query Match: 51.90% Indels: 0
DB: 11 Gaps: 0
US-09-737-297-3 (1-16) x US-09-864-408A-2561 (1-445)
QY 3 GlySerThr***AspValTyrglnAsnIleGlnTyrAla 15
DB 273 GGCAGCTGGGGTGATGCTATATAAAATCTGCAGTGGAGT 311
RESULT 13
US-09-815-242-7173/C
; Sequence 7173, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITEA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
```

```
US-09-737-297-3 (1-16) x US-09-815-242-7173 (1-1230)
QY 1 AlagluGlySerThr***AspValTyrglnAsnIleGlnTyr 14
DB 565 GCAACGGGTGCTATTGGCACACGTTTTTCAGAACGTGCGCTAT 524
RESULT 14
US-10-289-757-17
; Sequence 17, Application US/10289757
; Publication No. US20030180751A1
; GENERAL INFORMATION:
; APPLICANT: Demmer, Jerroen
; APPLICANT: Forster, Richard L.
; APPLICANT: Gibson, John Bryan
; APPLICANT: Shenk, Michael Andrew
; APPLICANT: No. US20030180751A1riss, Geoffrey
; APPLICANT: Glenn, Matthew
; APPLICANT: Saulsbury, Keith Martin
; APPLICANT: Hall, Claire
; TITLE OF INVENTION: Compositions isolated from forage
; TITLE OF INVENTION: grasses and methods for their use
; FILE REFERENCE: 11000.1061U
; CURRENT APPLICATION NUMBER: US/10/289,757
; CURRENT FILING DATE: 2002-11-07
; PRIOR APPLICATION NUMBER: 60/337,703
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 2010
; TYPE: DNA
; ORGANISM: Lolium perenne
US-10-289-757-17
Alignment Scores:
Pred. No.: 211 Length: 2010
Score: 41.00 Matches: 7
Percent Similarity: 69.23% Conservative: 2
Best Local Similarity: 53.85% Mismatches: 4
Query Match: 51.90% Indels: 0
DB: 14 Gaps: 0
US-09-737-297-3 (1-16) x US-10-289-757-17 (1-2010)
QY 2 GluGlySerThr***AspValTyrglnAsnIleGlnTyr 14
DB 929 GACAACCCCGCGCGAGCTGTACCGCGCTCCAGTAC 967
RESULT 15
US-10-114-170-206
; Sequence 206, Application US/10114170
; Publication No. US20030023075A1
; GENERAL INFORMATION:
```

APPLICANT: Blattner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/114,170
FILING DATE: 01-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/453,702
FILING DATE: 03-DEC-1999
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 206:
SEQUENCE CHARACTERISTICS:
LENGTH: 43360
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 206:
US-10-114-170-206
Alignment Scores:
Pred. No.: 9.23e+03 Length: 43360
Score: 41.00 Matches: 7
Percent Similarity: 62.50% Conservative: 3
Best Local Similarity: 43.75% Mismatches: 6
Query Match: 51.90% Indels: 0
DB: 14 Gaps: 0
US-09-737-297-3 (1-16) x US-10-114-170-206 (1-43360)
QY 1 AlaGluglySerThr***AspValTyrlnAsnIleGlnTyAlaGly 16
Db 30219 TCGTCAGGTTCCAACAGATACATTATCAGACTCTGCATATCCTGGC 30266
RESULT 16
US-10-114-170-261
Sequence 261, Application US/10114170
Publication No. US20030023075A1
GENERAL INFORMATION:
APPLICANT: Blattner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady

STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/114,170
FILING DATE: 01-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/453,702
FILING DATE: 03-DEC-1999
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 261:
SEQUENCE CHARACTERISTICS:
LENGTH: 45325 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 261:
US-10-114-170-261
Alignment Scores:
Pred. No.: 9.75e+03 Length: 45325
Score: 41.00 Matches: 7
Percent Similarity: 62.50% Conservative: 3
Best Local Similarity: 43.75% Mismatches: 6
Query Match: 51.90% Indels: 0
DB: 14 Gaps: 0
US-09-737-297-3 (1-16) x US-10-114-170-261 (1-45325)
QY 1 AlaGluglySerThr***AspValTyrlnAsnIleGlnTyAlaGly 16
Db 31121 TCGTCAGGTTCCAACAGATACATTATCAGACTCTGCATATCCTGGC 31168
RESULT 17
US-10-329-960-1
Sequence 1, Application US/10329960
Publication No. US20030099277A1
GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fra
FILE OF INVENTION: Thereof, and Uses Thereof
FILE REFERENCE: PB186PI
CURRENT APPLICATION NUMBER: US/10/329,960
CURRENT FILING DATE: 2003-01-02
PRIOR APPLICATION NUMBER: US 09/643,990
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: US 08/487,429
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: US 08/426,787
PRIOR FILING DATE: 1995-04-21
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1830121
TYPE: DNA
ORGANISM: Haemophilus influenzae

FEATURE:
NAME/KEY: misc feature
LOCATION: (4747)..(4747)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:

NAME/KEY: misc feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature

LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152530)..(152530)

Alignment Scores:

Pred. No.:	8.36e+05	Length:	1830121
Score:	41.00	Matches:	7
Percent Similarity:	71.43%	Conservative:	3
Best Local Similarity:	50.00%	Mismatches:	4
Query Match:	51.90%	Indels:	0
DB:	14	Gaps:	0

US-09-737-297-3 (1-16) x US-10-329-960-1 (1-1830121)

QY 1 AlaGluGlySerThr***AspValtyrGlnAsnIleGlnTyr 14
Db 1728229 GCRAACGGGTCTATTGGCACACGTTTTTCAGAACGTGGCTAT 1728270

RESULT 18

US-10-329-670-1
Sequence 1, Application US/10329670
Publication No. US20040018503A1
GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
FILE OF INVENTION: Thereof, and Uses Thereof
FILE REFERENCE: PB186P1
CURRENT APPLICATION NUMBER: US/10/329,670
CURRENT FILING DATE: 2002-12-24
PRIOR APPLICATION NUMBER: US 09/643,990
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: US 08/487,429
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: US 08/426,787
PRIOR FILING DATE: 1995-04-21
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1830121
TYPE: DNA
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: misc feature

LOCATION: (4747)..(4747)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (29238)..(29238)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (55369)..(55369)

FEATURE:					
NAME/KEY: misc feature					
LOCATION: (142750)..(142750)					
OTHER INFORMATION: n equals a, t, g or c					
FEATURE:					
NAME/KEY: misc feature					
LOCATION: (145058)..(145058)					
OTHER INFORMATION: n equals a, t, g or c					
FEATURE:					
NAME/KEY: misc feature					
LOCATION: (145171)..(145171)					
OTHER INFORMATION: n equals a, t, g or c					
FEATURE:					
NAME/KEY: misc feature					
LOCATION: (145942)..(145942)					
OTHER INFORMATION: n equals a, t, g or c					
FEATURE:					
NAME/KEY: misc feature					
LOCATION: (147197)..(147197)					
OTHER INFORMATION: n equals a, t, g or c					
FEATURE:					
NAME/KEY: misc feature					
LOCATION: (150841)..(150841)					
OTHER INFORMATION: n equals a, t, g or c					
FEATURE:					
NAME/KEY: misc feature					
LOCATION: (152500)..(152500)					
OTHER INFORMATION: n equals a, t, g or c					
FEATURE:					
NAME/KEY: misc feature					
LOCATION: (152530)..(152530)					
Alignment Scores:					
Pred. No.:	8.36e+05		Length:	1830121	
Score:	41.00		Matches:	7	
Percent Similarity:	71.43%		Conservative:	3	
Best Local Similarity:	50.00%		Mismatches:	0	
Query Match:	51.90%		Indels:	4	
DB:	15		Gaps:	0	
US-09-737-297-3 (1-16) x US-10-329-670-1 (1-1830121)					
OY	1	AAGUGlySeThr***AspValTyGlnAsnIleGlnTyr	14		
DB	1728229	GCAACGGTCATTCGCACACGTTTTTCAGACGTGGCTAT	1728270		
RESULT 19					
US-09-960-352-11946					
Sequence 11946, Application US/09960352					
Patent No. US20020137139A1					
GENERAL INFORMATION:					
APPLICANT: Warren, Wesley C.					
APPLICANT: Tao, Ningbing					
APPLICANT: Byatt, John C.					
APPLICANT: Mathalegan, Nagappan					
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH THE PREVENTION OF CANCER					
TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION					
FILE REFERENCE: 16511.006/37-21(10298)C					
CURRENT APPLICATION NUMBER: US/09/960.352					
CURRENT FILING DATE: 2001-09-24					
NUMBER OF SEQ ID NOS: 15112					
SEQ ID NO 11946					
LENGTH: 415					
TYPE: DNA					
ORGANISM: Bos taurus					
OTHER INFORMATION: Clone ID: 51-LTB3058-049-Q1-K1-B4					
US-09-960-352-11946					
Alignment Scores:					
Pred. No.:	48.9		Length:	415	
Score:	40.00		Matches:	6	
Percent Similarity:	71.43%		Conservative:	4	
Best Local Similarity:	42.86%		Mismatches:	4	

Query Match: 50.63% Indels: 0
DB: 9 Gaps: 0
US-09-737-297-3 (1-16) x US-09-860-352-11946 (1-415)
QY 2 GluGlySerThr***AspValTyrGlnAsnIleGlnTyrAla 15
Db 61 GAGAAACACCGTTCACATCTACCCCAATGTTCTATTCACGC 102

RESULT 20

US-09-815-242-6773
; Sequence 6773, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Cart, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 6773

LENGTH: 1371

TYPE: DNA

ORGANISM: Enterococcus faecalis

FEATURE:

NAME/KEY: CDS

LOCATION: (1)...(1371)

US-09-815-242-6773

Alignment Scores:
Pred. No.: 213 Length: 1371
Score: 40.00 Matches: 7
Percent Similarity: 60.00% Conservative: 2
Best Local Similarity: 46.67% Mismatches: 6
Query Match: 50.63% Indels: 0
DB: 9 Gaps: 0

US-09-737-297-3 (1-16) x US-09-815-242-6773 (1-1371)

QY 2 GluGlySerThr***AspValTyrGlnAsnIleGlnTyrAlagly 16
Db 763 GAAGGCTCATCTTTGATGTTATCAAGGCTGATTGTAGGT 807

RESULT 21

US-09-759-130B-182
; Sequence 182, Application US/09759130B
; Publication No. US2003002279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A

; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D
; APPLICANT: Barnes, Thomas S
; APPLICANT: Kirst, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodearl, Andrew
; APPLICANT: Holtzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; TITLE OF INVENTION: USES

FILE REFERENCE: MPI00-5350MNIM

CURRENT APPLICATION NUMBER: US/09/759,130B

CURRENT FILING DATE: 2002-09-16

PRIOR APPLICATION NUMBER: US 09/479,249

PRIOR FILING DATE: 2000-01-07

PRIOR APPLICATION NUMBER: US 09/559,497

PRIOR FILING DATE: 2000-04-27

PRIOR APPLICATION NUMBER: US 09/578,063

PRIOR FILING DATE: 2000-05-24

PRIOR APPLICATION NUMBER: US 09/333,159

PRIOR FILING DATE: 1999-06-14

PRIOR APPLICATION NUMBER: US 09/596,194

PRIOR FILING DATE: 2000-07-14

PRIOR APPLICATION NUMBER: US 09/342,364

PRIOR FILING DATE: 1999-06-29

PRIOR APPLICATION NUMBER: US 09/608,452

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/393,996

PRIOR FILING DATE: 1999-09-10

PRIOR APPLICATION NUMBER: US 09/602,871

PRIOR FILING DATE: 2000-06-23

PRIOR APPLICATION NUMBER: US 09/420,707

PRIOR FILING DATE: 1999-10-19

NUMBER OF SEQ ID NOS: 460

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 182

LENGTH: 1554

TYPE: DNA

ORGANISM: Mus sp.

US-09-759-130B-182

Alignment Scores:
Pred. No.: 248 Length: 1554
Score: 40.00 Matches: 7
Percent Similarity: 83.33% Conservative: 3
Best Local Similarity: 58.33% Mismatches: 2
Query Match: 50.63% Indels: 0
DB: 10 Gaps: 0

US-09-737-297-3 (1-16) x US-09-759-130B-182 (1-1554)

QY 2 GluGlySerThr***AspValTyrGlnAsnIleGln 13
Db 438 CAAGGTCACCTCCACTGACGTTTACCAGGATATCCAA 473

RESULT 22

US-10-276-629-7/c
; Sequence 7, Application US/10276629
; Publication No. US20030170261A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Christoph
; TITLE OF INVENTION: Virulence Genes, Proteins, and Their Use
; FILE REFERENCE: GJE-6443
; CURRENT APPLICATION NUMBER: US/10/276,629
; CURRENT FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/GB01/02247
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: GB 0012079.0
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 22


```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1773
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1773)
; US-10-276-629-7

Alignment Scores:
Pred. No.: 292 Length: 1773
Score: 40.00 Matches: 8
Percent Similarity: 62.50% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 50.63% Indels: 0
DB: 14 Gaps: 0

US-09-737-297-3 (1-16) x US-10-276-629-7 (1-1773)

Qy 1 AlaGluGlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16
Db 1506 GCGTTGGCGGTACGATACGCTGTTCAGCATTTTCAGCAGCGGG 1459

RESULT 23
; Sequence 29515, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITEA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29515
; LENGTH: 1776
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
; US-10-282-122A-29515

US-09-737-297-3 (1-16) x US-10-282-122A-29515 (1-1776)

Qy 1 AlaGluGlySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16
Db 1506 GCGTTGGCGGTACGATACGCTGTTCAGCATTTTCAGCAGCGGG 1459

RESULT 24
; Sequence 644, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 644
; LENGTH: 1793
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (790)
; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: misc feature
; LOCATION: (1731)
; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: misc feature
; LOCATION: (1793)
; OTHER INFORMATION: n equals a.t.g, or c
; US-09-925-300-644

Alignment Scores:
Pred. No.: 296 Length: 1793
Score: 40.00 Matches: 6
Percent Similarity: 90.91% Conservative: 4
Best Local Similarity: 54.55% Mismatches: 1
Query Match: 50.63% Indels: 0
DB: 9 Gaps: 0

US-09-737-297-3 (1-16) x US-09-925-300-644 (1-1793)

Qy 5 Thr***AspValTyrGlnAsnIleGlnTyrAla 15
Db 1668 ACTCAAGATATTTATAGATCAGTTCATATGCT 1636

RESULT 25
; Sequence 28, Application US/10273051
; Publication No. US20030157673A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Mario W.
; APPLICANT: Chen, Qiong
; APPLICANT: Gibson, Katherine J.
; APPLICANT: Kostichka, Kristy N.
; APPLICANT: Thomas, Stuart M.
; APPLICANT: NagaraJan, Vasantha
```

```

; TITLE OF INVENTION: Genes Involved in Cyclododecanone Degradation Pathway
; FILE REFERENCE: BC1023 US NA
; CURRENT APPLICATION NUMBER: US/10/273,051
; CURRENT FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: 60/170,214
; PRIOR FILING DATE: 1999-12-10
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 28
; LENGTH: 1812
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Cdda Homolog
US-10-273-051-28

Alignment Scores:
Pred. No.: 300 Length: 1812
Score: 40.00 Matches: 7
Percent Similarity: 68.75% Conservative: 4
Best Local Similarity: 43.75% Mismatches: 5
Query Match: 50.63% Indels: 0
DB: 14 Gaps: 0

US-09-737-297-3 (1-16) x US-10-273-051-28 (1-1812)
Qy 1 AlaGluGlySerThr***AspValTyrglnAsnIleGlnTyAlaGly 16
Db 806 GCCGAGGCGGTGAGTGGACTACGTACCGATCCCTGCAGTACGTGGGC 759

RESULT 26
US-10-129-518-28/c
; Sequence 28, Application US/10129518
; Publication No. US20030215930A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Company
; TITLE OF INVENTION: Genes Involved in Cyclododecanone Degradation Pathway
; FILE REFERENCE: BC1023 PCT
; CURRENT APPLICATION NUMBER: US/10/129,518
; CURRENT FILING DATE: 2002-11-04
; PRIOR APPLICATION NUMBER: 60/170,214
; PRIOR FILING DATE: 1999-12-10
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 28
; LENGTH: 1812
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Cdda Homolog
US-10-129-518-28

Alignment Scores:
Pred. No.: 300 Length: 1812
Score: 40.00 Matches: 7
Percent Similarity: 68.75% Conservative: 4
Best Local Similarity: 43.75% Mismatches: 5
Query Match: 50.63% Indels: 0
DB: 15 Gaps: 0

US-09-737-297-3 (1-16) x US-10-129-518-28 (1-1812)
Qy 1 AlaGluGlySerThr***AspValTyrglnAsnIleGlnTyAlaGly 16
Db 806 GCCGAGGCGGTGAGTGGACTACGTACCGATCCCTGCAGTACGTGGGC 759

RESULT 27
US-09-845-157-1
; Sequence 1, Application US/09845157
; Patent No. US20020090618A1
; GENERAL INFORMATION:
; APPLICANT: Smith, J.
; TITLE OF INVENTION: Thermostable Reverse Transcriptases and Uses Thereof
```

```

; FILE REFERENCE: 0942.5040001
; CURRENT APPLICATION NUMBER: US/09/845,157
; CURRENT FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: US 60/207,196
; PRIOR FILING DATE: 2000-03-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: M-MLV reverse transcriptase gene
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2151)
US-09-845-157-1

Alignment Scores:
Pred. No.: 370 Length: 2151
Score: 40.00 Matches: 8
Percent Similarity: 66.67% Conservative: 2
Best Local Similarity: 53.33% Mismatches: 5
Query Match: 50.63% Indels: 0
DB: 9 Gaps: 0

US-09-737-297-3 (1-16) x US-09-845-157-1 (1-2151)
Qy 1 AlaGluGlySerThr***AspValTyrglnAsnIleGlnTyAla 15
Db 1816 GCAGAGGTAGAGCTAATGTTATACGAATCCCGTTATGCT 1860

RESULT 28
US-09-759-130B-181
; Sequence 181, Application US/09759130B
; Publication No. US2003022279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A
; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D
; APPLICANT: Barnes, Thomas S
; APPLICANT: Kirst, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodearl, Andrew
; APPLICANT: Holtzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; TITLE OF INVENTION: USES.
; FILE REFERENCE: MP100-5350NMIM
; CURRENT APPLICATION NUMBER: US/09/759,130B
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
```

```

; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 181
; LENGTH: 2467
; TYPE: DNA
; ORGANISM: Mus sp.
US-09-759-130B-181

Alignment Scores:
438      Length: 2467
40.00    Matches: 7
83.33%   Conservative: 3
58.33%   Mismatches: 2
50.63%   Indels: 0
10        Gaps: 0

US-09-737-297-3 (1-16) x US-09-759-130B-181 (1-2467)
QY      2 GluGlySerThr***AspValTyGlnAsnIleGln 13
      :::::|||||:::|||||
Ddb     438 CAAGGTCACTCCACTGACGCTTTACCAGGATATCCAA 473

RESULT 29
US-10-044-090-796/c
; Sequence 796, Application US/10044090
; Publication No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 796
; LENGTH: 3691
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 234196.3
US-10-044-090-796

Alignment Scores:
719      Length: 3691
40.00    Matches: 6
90.91%   Conservative: 4
54.55%   Mismatches: 1
50.63%   Indels: 0
13        Gaps: 0

US-09-737-297-3 (1-16) x US-10-044-090-796 (1-3691)
QY      5 Thr***AspValTyGlnAsnIleGlnTyAla 15
      |||:::|||||:::|||||
Ddb     1784 ACTGAGATATTATTATAGATCACTTCATATGCT 1752

RESULT 30
US-10-027-632-76212
; Sequence 76212, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20

```

```

; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 235:
US-10-158-844-235

Alignment Scores:
Pred. No.: 369 Length: 1766
Score: 39.50 Matches: 9
Percent Similarity: 55.00% Conservative: 2
Best Local Similarity: 45.00% Mismatches: 4
Query Match: 50.00% Indels: 5
DB: 12 Gaps: 1

US-09-737-297-3 (1-16) x US-10-158-844-235 (1-1766)

OY 2 GluglySerThr***AspValTyrGlnAsn-----IleGlnTyrAlaGly 16
Db 190 GAAGGAGAAATCGATTGTGATCAATAAAGGCTGTTTTATACAAATATGCTGT 249

RESULT 32
US-10-131-827-6245/c
; Sequence 6245, Application US/10131827
; Publication No. US20040009479A1
; GENERAL INFORMATION:
; APPLICANT: Wohlgemuth, Jay
; APPLICANT: Fry, Kirk, Robert
; APPLICANT: Woodward, Robert
; TITLE OF INVENTION: LY, NGOC
; TITLE OF INVENTION: CHRONIC INFLAMMATORY DISEASES
; FILE REFERENCE: 506612000120
; CURRENT APPLICATION NUMBER: US/10/131,827
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US 10/006,290
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/296,764
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 9090
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6245
; LENGTH: 50
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-131-827-6245

Alignment Scores:
Pred. No.: 584 Length: 50
Score: 39.00 Matches: 7
Percent Similarity: 80.00% Conservative: 1
Best Local Similarity: 70.00% Mismatches: 2
Query Match: 49.37% Indels: 0
DB: 15 Gaps: 0

US-09-737-297-3 (1-16) x US-10-131-827-6245 (1-50)

OY 1 AlaGluglySerThr***AspValTyrGln 10
Db 46 GCACAGGAGAAACCCACAGACGCTTTATCAG 17

RESULT 33
US-10-027-632-87366/c
; Sequence 87366, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 138455
; LENGTH: 847
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-138455

Alignment Scores:
Pred. No.: 190 Length: 847
Score: 39.00 Matches: 7
Percent Similarity: 68.75% Conservative: 4

```

```

; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 87366
; LENGTH: 563
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)-(563)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-87366

Alignment Scores:
Pred. No.: 115 Length: 563
Score: 39.00 Matches: 7
Percent Similarity: 60.00% Conservative: 2
Best Local Similarity: 46.67% Mismatches: 6
Query Match: 49.37% Indels: 0
DB: 15 Gaps: 0

US-09-737-297-3 (1-16) x US-10-027-632-87366 (1-563)

OY 2 GluglySerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16
Db 409 GAGGGATCCAGGAACTACCTACATCGACTTCACTTCTGCT 365

RESULT 34
US-10-027-632-138455/c
; Sequence 138455, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 138455
; LENGTH: 847
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-138455

Alignment Scores:
Pred. No.: 190 Length: 847
Score: 39.00 Matches: 7
Percent Similarity: 68.75% Conservative: 4

```

Best Local Similarity: 43.75% Mismatches: 5
Query Match: 49.37% Indels: 0
DB: 15 Gaps: 0

US-09-737-297-3 (1-16) x US-10-027-632-138455 (1-847)

QY 1 AlaGlucylSerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16
DB 780 GCAGAGGGGAATCACTAACACATTTTAAATAATCTCTCCAGGCTGGG 733

RESULT 35
US-10-027-632-157309/c
Sequence 157309, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: Fast-Seq for Windows Version 4.0
SEQ ID NO 157309
LENGTH: 847
TYPE: DNA
ORGANISM: Human
US-10-027-632-157309

Alignment Scores:
Pred. No.: 190 Length: 847
Score: 39.00 Matches: 7
Percent Similarity: 68.75% Conservative: 4
Best Local Similarity: 43.75% Mismatches: 5
Query Match: 49.37% Indels: 0
DB: 15 Gaps: 0

US-09-737-297-3 (1-16) x US-10-027-632-157309 (1-847)

QY 1 AlaGlucylSerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16
DB 780 GCAGAGGGGAATCACTAACACATTTTAAATAATCTCTCCAGGCTGGG 733

RESULT 36
US-10-369-493-32618
Sequence 32618, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 32618
LENGTH: 1320
TYPE: DNA
ORGANISM: Chloroflexus aurantiacus
US-10-369-493-32618

Alignment Scores:
Pred. No.: 327 Length: 1320
Score: 39.00 Matches: 7
Percent Similarity: 69.23% Conservative: 2
Best Local Similarity: 53.85% Mismatches: 4
Query Match: 49.37% Indels: 0
DB: 15 Gaps: 0

US-09-737-297-3 (1-16) x US-10-369-493-32618 (1-1320)

QY 4 SerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16
DB 427 AACACGGAGATGTTTACACCGCATCGAGTATCGCGGC 465

RESULT 37
US-10-282-122A-24198
Sequence 24198, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsen
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITEA-034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: Patent in version 3.1
SEQ ID NO 24198
LENGTH: 1872
TYPE: DNA
ORGANISM: Listeria monocytogenes
US-10-282-122A-24198

Alignment Scores:

Alignment Scores: 503 1872
Pred. No.: 39.00 6
Score: 75.00% Matches: 3
Percent Similarity: 35.00% Conservative: 3
Best Local Similarity: 50.00% Mismatches: 3
Query Match: 49.37% Indels: 0
DB: 12 Gaps: 0

US-09-737-297-3 (1-16) x US-10-282-122A-24198 (1-1872)

2y 2 GlucySerThr***AspValTyrGlnAsnIleGln 13
Db 175 GATGGATCTACACAGATACATACGACGTTTGCAA 210

RESULT 38
US-10-398-221-847
; Sequence 847, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; PRIOR FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 847
; LENGTH: 1872
; TYPE: DNA
; ORGANISM: Listeria monocytogenes-EGD
US-10-398-221-847

Alignment Scores: 503 1872
Pred. No.: 39.00 6
Score: 75.00% Matches: 3
Percent Similarity: 35.00% Conservative: 3
Best Local Similarity: 50.00% Mismatches: 3
Query Match: 49.37% Indels: 0
DB: 15 Gaps: 0

US-09-737-297-3 (1-16) x US-10-398-221-847 (1-1872)

2y 2 GlucySerThr***AspValTyrGlnAsnIleGln 13
Db 175 GATGGATCTACACAGATACATACGACGTTTGCAA 210

RESULT 39
US-10-398-221-2762
; Sequence 2762, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; PRIOR FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2762
; LENGTH: 1872
; TYPE: DNA
; ORGANISM: Listeria monocytogenes EGDe
US-10-398-221-2762

Alignment Scores: 503 1872
Pred. No.: 39.00 6
Score: 75.00% Matches: 3
Percent Similarity: 35.00% Conservative: 3
Best Local Similarity: 50.00% Mismatches: 3
Query Match: 49.37% Indels: 0
DB: 15 Gaps: 0

US-09-737-297-3 (1-16) x US-10-398-221-2762 (1-1872)

2y 2 GlucySerThr***AspValTyrGlnAsnIleGln 13
Db 175 GATGGATCTACACAGATACATACGACGTTTGCAA 210

RESULT 40
US-09-887-576-118/c
; Sequence 118, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360 001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 118
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-887-576-118

Alignment Scores: 546 2000
Pred. No.: 39.00 7
Score: 69.23% Matches: 2
Percent Similarity: 53.85% Conservative: 4
Best Local Similarity: 49.37% Indels: 0
Query Match: 9 Gaps: 0
DB:

US-09-737-297-3 (1-16) x US-09-887-576-118 (1-2000)

2y 3 GlySerThr***AspValTyrGlnAsnIleGlnTyrAla 15
Db 536 GGAACCCACGACGATCTTACCAAAACCTCAGTACGCG 498

RESULT 41
US-09-938-842A-3048
; Sequence 3048, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SAME, AND METHODS OF USE
; TITLE OF INVENTION: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866

DB: 15 Gaps: 0

US-09-737-297-3 (1-16) x US-10-264-049-205 (1-3907)

QY 4 SerThr***AspValTyrGlnAsnIleGlnTyrAlaGly 16

DB 2664 TCAAAATTCAGATATTATATGATATGATGAATACCATGGT 2702

RESULT 48

US-09-764-847-1514/c

; Sequence 1514, Application US/09764847

; Patent No. US20020132767A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC009

; CURRENT APPLICATION NUMBER: US/09/764,847

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 2003

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1514

; LENGTH: 23378

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-764-847-1514

Alignment Scores:

Pred. No.: 1.12e+04 Length: 23378

Score: 39.00 Matches: 6

Percent Similarity: 78.57% Conservative: 5

Best Local Similarity: 42.86% Mismatches: 3

Query Match: 49.37% Indels: 0

DB: 9 Gaps: 0

US-09-737-297-3 (1-16) x US-09-764-847-1514 (1-23378)

QY 1 AlaGluGlySerThr***AspValTyrGlnAsnIleGlnTyr 14

DB 7876 GCTAAAGGAGCCTCAGTGATATGATTTCAAAACTTGGATGG 7835

RESULT 49

US-10-092-154-1514/c

; Sequence 1514, Application US/10092154

; Publication No. US20030054375A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC009C1

; CURRENT APPLICATION NUMBER: US/10/092,154

; CURRENT FILING DATE: 2002-03-07

; NUMBER OF SEQ ID NOS: 2003

; Prior Application removed - See File Wrapper or Palm

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1514

; LENGTH: 23378

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-092-154-1514

Alignment Scores:

Pred. No.: 1.12e+04 Length: 23378

Score: 39.00 Matches: 6

Percent Similarity: 78.57% Conservative: 5

Best Local Similarity: 42.86% Mismatches: 3

Query Match: 49.37% Indels: 0

DB: 14 Gaps: 0

US-09-737-297-3 (1-16) x US-10-092-154-1514 (1-23378)

QY 1 AlaGluGlySerThr***AspValTyrGlnAsnIleGlnTyr 14

DB 7876 GCTAAAGGAGCCTCAGTGATATGATTTCAAAACTTGGATGG 7835

RESULT 50

US-09-731-231A-3/c

; Sequence 3, Application US/09731231A

; Patent No. US20020082189A1

; GENERAL INFORMATION:

; APPLICANT: GURGLER, Karl et al

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; FILE REFERENCE: THEREOF

; CURRENT APPLICATION NUMBER: US/09/731,231A

; CURRENT FILING DATE: 2000-12-07

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 326014

; TYPE: DNA

; ORGANISM: Human

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)..(326014)

; OTHER INFORMATION: n = A, T, C or G

US-09-731-231A-3

Alignment Scores:

Pred. No.: 2.78e+05 Length: 326014

Score: 39.00 Matches: 7

Percent Similarity: 75.00% Conservative: 2

Best Local Similarity: 58.33% Mismatches: 3

Query Match: 49.37% Indels: 0

DB: 9 Gaps: 0

US-09-737-297-3 (1-16) x US-09-731-231A-3 (1-326014)

QY 5 Thr***AspValTyrGlnAsnIleGlnTyrAlaGly 16

DB 123595 ACATGGACATTTCACCAAAATTGCAATATACAGG 123560

Search completed: March 1, 2004, 12:03:29

Job time : 1703 secs